

67655

STIC-Biotech/ChemLib

From: Davis, Natalie
Sent: Thursday, May 30, 2002 10:23 AM
To: STIC-Biotech/ChemLib
Subject: 09/589777

Please search SEQ ID NO: 24-25 and interference for 09/589777

Thanks

Natalie A. Davis, PhD
Patent Examiner
Art Unit 1642
CM1, Rm 8B13
Mailbox 8E12
Ph (703) 308-6410

CRFE

if Contact:
Sheppard

Searcher: _____
Phone: tel: 308-4499
Location: _____
Date Picked Up: _____
Date Completed: 6/3/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: May 31, 2002, 10:27:24 ; Search time 41.34 Seconds
(without alignments)
33.477 Million cell updates/sec

Title: US-09-589-777C-24
Perfect score: 41
Sequence: 1 SYIVLCIE 8
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhct:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match %	Length	ID	Description
1	41	100.0	160	11 Q9CRT2	Q9crt2 mus musculu
2	41	100.0	171	11 Q9WUW5	Q9wuw5 rattus norv
3	41	100.0	184	11 Q9JK63	Q9jk63 mus musculu
4	41	100.0	226	11 Q9QZD2	Q9qzd2 rattus norv
5	41	100.0	1140	11 Q61434	Q61434 mus musculu
6	41	100.0	1774	11 Q62001	Q62001 mus musculu
7	33	80.5	237	2 O87543	O87543 capnocytoph
8	33	80.5	646	5 O15900	O15900 dictyosteli
9	33	80.5	706	5 O15901	O15901 dictyosteli
10	33	80.5	1344	13 Q93419	Q93419 gallus gall
11	33	80.5	1905	5 Q9XTP6	Q9xtp6 plasmodium
12	32	78.0	147	16 O84569	O84569 chlamydia t
13	32	78.0	147	16 Q9PJH5	Q9pjh5 chlamydia m
14	32	78.0	215	16 Q97TL1	Q97tl1 clostridium
15	32	78.0	254	10 Q9C7K8	Q9c7k8 arabidopsis
16	32	78.0	293	10 Q9C7Y0	Q9c7y0 arabidopsis

17	32	78.0	347	1 Q9P9G6	Q9p9g6 methanococc
18	32	78.0	507	3 Q9P905	Q9p905 emericella
19	32	78.0	613	4 Q95942	Q95942 homo sapien
20	32	78.0	913	2 Q93M96	Q93m96 clostridium
21	31	75.6	66	16 Q932K0	Q932k0 staphylococ
22	31	75.6	70	4 Q9P1N2	Q9pln2 homo sapien
23	31	75.6	152	13 Q90W88	Q90w88 bombina max
24	31	75.6	158	16 Q9Z784	Q9z784 chlamydia p
25	31	75.6	163	10 Q9SLC0	Q9slc0 arabidopsis
26	31	75.6	174	16 Q97HJ6	Q97hj6 clostridium
27	31	75.6	201	10 Q9AVY4	Q9avy4 guillardia
28	31	75.6	208	13 Q90WZ1	Q90wz1 bombina max
29	31	75.6	213	12 Q9J5B0	Q9j5b0 fowlpox vir
30	31	75.6	235	5 Q93866	Q93866 caenorhabdi
31	31	75.6	272	12 Q9EN36	Q9en36 amsacta moo
32	31	75.6	316	16 Q9RST2	Q9rst2 deinococcus
33	31	75.6	466	16 Q9KMS0	Q9kms0 vibrio chol
34	31	75.6	472	5 Q9VQN6	Q9vqn6 drosophila
35	31	75.6	479	10 Q9XI52	Q9xi52 arabidopsis
36	31	75.6	566	16 Q9PHU2	Q9phu2 campylobact
37	31	75.6	861	10 Q9AVX8	Q9avx8 guillardia
38	31	75.6	991	5 Q95024	Q95024 dictyosteli
39	31	75.6	1602	10 Q94E43	Q94e43 oryza sativ
40	31	75.6	2158	10 Q9LUT5	Q9lut5 arabidopsis
41	31	75.6	2342	5 O01677	O01677 bombyx mori
42	30	73.2	33	5 Q95S45	Q95s45 drosophila
43	30	73.2	96	9 O48440	O48440 bacterioph
44	30	73.2	105	2 Q93MC5	Q93mc5 clostridium
45	30	73.2	129	5 O62211	O62211 caenorhabdi

ALIGNMENTS

RESULT 1
Q9CRT2
ID Q9CRT2 PRELIMINARY; PRT; 160 AA.
AC Q9CRT2;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE PROCOLLAGEN, TYPE XVIII, ALPHA 1 (FRAGMENT).
GN COL18A1.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]

SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shiragawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Tovo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK014292; BAB29249.1; -.
DR HSSP; P39061; IKOE.

DR MGD; MGI:88451; Coll18a1.
FT NON_TER 1 1
SQ SEQUENCE 160 AA; 17725 MW; 60F853D777C7375D2 CRC64;

Query Match 100.0%; Score 41; DB 11; Length 160;
Best Local Similarity 100.0%; Pred. No. 0.41; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 144 SYIVLCIE 151

RESULT 2
Q9WUW5 PRELIMINARY; PRT; 171 AA.
ID Q9WUW5
AC Q9WUW5;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE COLLAGEN TYPE XVIII, ALPHA (I) CHAIN (FRAGMENT).
GN COL18A1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Jia J.D., Bauer M., Eberspaecher U., Donner P., Schuppan D.;
RT "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Jia J.D., Bauer M., Sedlacek N., Ruehl M., Riecken E.O., Schuppan D.;
RT "Temporospatial expression of collagen XVIII/endostatin in acute and chronic liver injuries";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ236873; CAB44263.1; -
DR HSSP; P39061; IKOE.
FT NON_TER 1 1
FT NON_TER 171 171
SQ SEQUENCE 171 AA; 18933 MW; 81BE2EE3FC2C8E72 CRC64;

Query Match 100.0%; Score 41; DB 11; Length 171;
Best Local Similarity 100.0%; Pred. No. 0.44; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 160 SYIVLCIE 167

RESULT 3
Q9JK63 PRELIMINARY; PRT; 184 AA.
ID Q9JK63
AC Q9JK63;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ENDOSTATIN (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHINESE KUNMING;
RA Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;
RT "Anticancer treatment of targeted fusion protein delivery to tumor

RT neovasculature";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF257775; AAF69009.1; -
DR HSSP; P39061; IKOE.
FT NON_TER 1 1
FT NON_TER 184 184
SQ SEQUENCE 184 AA; 20376 MW; AC06F9D8D103412A CRC64;

Query Match 100.0%; Score 41; DB 11; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.47; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 168 SYIVLCIE 175

RESULT 4
Q9QZD2 PRELIMINARY; PRT; 226 AA.
ID Q9QZD2
AC Q9QZD2;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE COLLAGEN XVIII (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=20227226; PubMed=10766159;
RA Perletti G., Concarl P., Giardini R., Marras E., Piccinini F.,
RA Folkman J., Chen L.;
RT "Antitumor activity of endostatin against carcinogen-induced rat primary mammary tumors";
RL Cancer Res. 60:1793-1796(2000).
DR EMBL; AF189709; AAF00975.1; -
DR HSSP; P39061; IKOE.
FT NON_TER 1 1
FT NON_TER 226 226
SQ SEQUENCE 226 AA; 25350 MW; 38B83C0486C0E949 CRC64;

Query Match 100.0%; Score 41; DB 11; Length 226;
Best Local Similarity 100.0%; Pred. No. 0.57; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 210 SYIVLCIE 217

RESULT 5
Q61434 PRELIMINARY; PRT; 1140 AA.
ID Q61434
AC Q61434;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE COLLAGEN (FRAGMENT).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe N., Muragaki Y., Yoshioka H., Inoue H., Ninomiya Y.;
RT "Identification of a novel collagen chain represented by extensive interruptions in the triple-helical region.";
RL Cell. Mol. Biol. Res. 196:576-582(1993).

DR EMBL; D17546; BAA04483.1; -
DR HSSP; P39061; 1KOE.
DR MGD; MGI:88449; Coll15a1.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 6.
FT NON_TER 1
SQ SEQUENCE 1140 AA; 115156 MW; 8BOC7E5862B3BDFE CRC64;

Query Match 100.0%; Score 41; DB 11; Length 1140;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 1124 SYIVLCIE 1131
|||||

RESULT 6
Q62001 PRELIMINARY; PRT; 1774 AA.
ID Q62001; Q60672;
AC Q62001; Q60672;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
DE PROCOLLAGEN, TYPE XVIII, ALPHA 1 PRECURSOR (XVIII) COLLAGEN
DE (PROCOLLAGEN, TYPE XVIII, ALPHA 1) (ALPHA-1 TYPE XVIII COLLAGEN).
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PRIMARY TAIL CULTURE;
RX MEDLINE=94245707; PubMed=8188673;
RA Rehn M., Hintikka E., Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT chain.";
RL J. Biol. Chem. 269:13929-13935(1994).
RN [2]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=94240112; PubMed=8183894;
RA Rehn M., Pihlajaniemi T.;
RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN [3]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=95181468; PubMed=7876242;
RA Rehn M., Pihlajaniemi T.;
RT "Identification of three N-terminal ends of type XVIII collagen chains
RT and tissue-specific differences in the expression of the corresponding
RT transcripts. The longest form contains a novel motif homologous to rat
RT and Drosophila frizzled proteins.";
RL J. Biol. Chem. 270:4705-4711(1995).
DR EMBL; U03715; AAC52903.1; -
DR EMBL; U03716; AAC52903.1; JOINED.
DR EMBL; U03718; AAC52903.1; JOINED.
DR EMBL; U34607; AAC52903.1; JOINED.
DR EMBL; U34608; AAC52903.1; JOINED.
DR EMBL; U34609; AAC52903.1; JOINED.
DR EMBL; U34610; AAC52903.1; JOINED.
DR EMBL; U34611; AAC52903.1; JOINED.
DR EMBL; U34612; AAC52903.1; JOINED.
DR EMBL; U34613; AAC52903.1; JOINED.
DR EMBL; U11637; AAC52179.1; -
DR HSSP; P39061; 1KOE.
DR MGD; MGI:88451; Coll18a1.
DR InterPro; IPR000087; Collagen.

DR InterPro; IPR000024; Fz_domain.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 6.
DR Pfam; PF01392; Fz; 1.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00063; FRI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS50038; FZ; 1.
KW Signal.
SQ SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88EF232 CRC64;

Query Match 100.0%; Score 41; DB 11; Length 1774;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 1758 SYIVLCIE 1765
|||||

RESULT 7
O87543 PRELIMINARY; PRT; 237 AA.
ID O87543
AC O87543;
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN ADHESIN (FRAGMENT).
OS Capnocytophaga gingivalis.
OC Bacteria; CFB group; Flavobacteriia; Flavobacteriaceae; Capnocytophaga.
OX NCBI_TaxID=1017;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DR2001;
RA Tempro P.J.;
RT "Capnocytophaga outer-membrane adhesin DNA sequence, 3'end.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083393; AAC35353.1; -
DR MEROPS; S09.013; -
DR InterPro; IPR000379; Est_lip_thioest_actsite.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Prol_endopep_ser.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
FT NON_TER 1
SQ SEQUENCE 237 AA; 27482 MW; 4D630AC81D64EDAC CRC64;

Query Match 80.5%; Score 33; DB 2; Length 237;
Best Local Similarity 71.4%; Pred. No. 29;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 YIVLCIE 8
Db 62 YIVLCVD 68
|||||

RESULT 8
O15900 PRELIMINARY; PRT; 646 AA.
ID O15900
AC O15900;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE PUTATIVE TRANSCRIPTION INHIBITOR.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.

RC TRANSPOSON=TDD-4;
RX MEDLINE=99263047; PubMed=10325432;
RA Wells D.J.;
RT "Tdd-4, a DNA transposon of Dictyostelium that encodes proteins
similar to LTR retroelement integrases.";
RL Nucleic Acids Res. 27:2408-2415(1999).
DR EMBL; U57081; AAB95435.1; -.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00665; rve; 1.
SQ SEQUENCE 646 AA; 74364 MW; 89E4F091505B2D4E CRC64;

Query Match 80.5%; Score 33; DB 5; Length 646;
Best Local Similarity 71.4%; Pred. No. 74;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 YIVLCIE 8
||:||||
Db 185 YIILCID 191

RESULT 9
O15901
ID O15901 PRELIMINARY; PRT; 706 AA.
AC O15901;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PUTATIVE TRANSPOSASE.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=TDD-4;
RX MEDLINE=99263047; PubMed=10325432;
RA Wells D.J.;
RT "Tdd-4, a DNA transposon of Dictyostelium that encodes proteins
similar to LTR retroelement integrases.";
RL Nucleic Acids Res. 27:2408-2415(1999).
DR EMBL; U57081; AAB95436.1; -.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00665; rve; 1.
SQ SEQUENCE 706 AA; 80670 MW; F5718916484A3EE5 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 706;
Best Local Similarity 71.4%; Pred. No. 80;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 YIVLCIE 8
||:||||
Db 185 YIILCID 191

RESULT 10
O93419
ID O93419 PRELIMINARY; PRT; 1344 AA.
AC O93419;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE COLLAGEN XVIII PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411346; PubMed=9738008;
RA Halfter W., Dong S., Schurer B., Cole G.J.;
RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.";

RL J. Biol. Chem. 273:25404-25412(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Halfter W., Dong S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083440; AAC33294.2; -.
DR HSSP; P39061; 1KOE.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 7.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Signal.
FT SIGNAL 1 26 POTENTIAL.
SQ SEQUENCE 1344 AA; 137402 MW; 7AA366E4FE940CCD CRC64;

Query Match 80.5%; Score 33; DB 13; Length 1344;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
:::|||||
Db 1328 AFVLCIE 1335

RESULT 11
Q9XTP6
ID Q9XTP6 PRELIMINARY; PRT; 1905 AA.
AC Q9XTP6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CTRP PROTEIN PRECURSOR (OOKINETE PROTEIN).
GN CTRP.
OS Plasmodium berghei (strain Anka).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ANKA;
RA Yuda M., Sawai T., Chinzei Y.;
RT "Structure and expression of an adhesive protein-like molecule of
mosquito invasive-stage malaria parasite.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ANKA;
RA Yuda M., Sawai T., Chinzei Y.;
RT "Structure and Expression of an Adhesive Protein-like Molecule of
Mosquito Invasive-stage Malarial Parasite.";
RL J. Exp. Med. 0:0-0(1999).
DR EMBL; AJ238798; CAB45562.1; -.
DR EMBL; AB027129; BAA82322.1; -.
DR EMBL; AF149771; AAF73158.1; -.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00090; tsp_1; 4.
DR Pfam; PF00092; vwa; 5.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00209; TSP1; 5.
DR SMART; SM00327; VWA; 6.
DR PROSITE; PS50092; TSP1; 4.
DR PROSITE; PS50234; VWFA; 8.

KW Signal. 1 21 POTENTIAL.
FT SIGNAL 22 1905 CTRP PROTEIN.
FT CHAIN 22 1905 CTRP PROTEIN.
SQ SEQUENCE 1905 AA; 215138 MW; 27A94B778CABDC36 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 1905;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
|:|||||
Db 10 SYLVLCV 16

RESULT 12
O84569
ID O84569 PRELIMINARY; PRT; 147 AA.
AC O84569;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 16.0 KDA PROTEIN.
GN CT565.

OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;

RA "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
DR EMBL; AE001327; AAC68167.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 147 AA; 15980 MW; 19D0EE2CE2F9DE4E CRC64;

Query Match 78.0%; Score 32; DB 16; Length 147;
Best Local Similarity 85.7%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
|:|||||
Db 123 SFIVLCI 129

RESULT 13
Q9PJH5
ID Q9PJH5 PRELIMINARY; PRT; 147 AA.
AC Q9PJH5;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN TC0854.
GN TC0854.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOBN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;

RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia

RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AE002352; AAF39650.1; -.
DR TIGR; TC0854; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 147 AA; 15871 MW; 6B33E9C475F8FD66 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 147;
Best Local Similarity 85.7%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
|:|||||
Db 123 SFIVLCI 129

RESULT 14
Q97TL1
ID Q97TL1 PRELIMINARY; PRT; 215 AA.
AC Q97TL1;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE PREDICTED MEMBRANE PROTEIN.
GN CAP0089.

OS Clostridium acetobutylicum.
OG Plasmid pSOLL.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;

RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE001438; AAK76835.1; -.
KW Plasmid; Complete proteome.
SQ SEQUENCE 215 AA; 25137 MW; 96DD09BDIFA9ACF1 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 215;
Best Local Similarity 71.4%; Pred. No. 42;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 YIVLCIE 8
|:|:||||
Db 77 YLVVCIE 83

RESULT 15
Q9C7K8
ID Q9C7K8 PRELIMINARY; PRT; 254 AA.
AC Q9C7K8;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE RNA PSEUDOURIDYLATE SYNTHASE, PUTATIVE.
GN FL4G9.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;

RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etnu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
RL Nature 408:816-820(2000).
DR EMBL; AC069159; AAG50905.1; -.
DR InterPro; IPR000613; Pseudou_synth.
DR InterPro; IPR002990; PSI_RLU.
DR Pfam; PF00849; Pseudou_synth_2; 1.
DR ProDom; PD001819; Pseudou_synth; 1.
DR PROSITE; PS01129; PSI_RLU; UNKNOWN_1.
SQ SEQUENCE 254 AA; 28236 MW; 0575EFC7E6439C93 CRC64;

Query Match 78.0%; Score 32; DB 10; Length 254;
Best Local Similarity 85.7%; Pred. No. 49;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
||| |||
Db 83 SYIALCI 89

Search completed: May 31, 2002, 10:33:06
Job time: 342 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:27:54 ; Search time 13.58 Seconds
(without alignments)
22.810 Million cell updates/sec

Title: US-09-589-777C-24
Perfect score: 41
Sequence: 1 SYIVLCIE 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	41	100.0	1527	1	CALH_MOUSE P39061 mus musculus
2	38	92.7	1516	1	CALH_HUMAN P39060 homo sapien
3	32	78.0	1015	1	YAL17_HUMAN Q9upz3 homo sapien
4	32	78.0	1142	1	GLG1_CHICK Q02391 gallus gall
5	32	78.0	1160	1	GLG1_CRIGR Q9z1e9 cricetus
6	32	78.0	1171	1	GLG1_RAT Q62638 rattus norv
7	32	78.0	1175	1	GLG1_MOUSE Q61543 mus musculus
8	32	78.0	1179	1	GLG1_HUMAN Q92896 homo sapien
9	30	73.2	233	1	YA36_METJA Q58442 methanococc
10	30	73.2	346	1	RDS_FELCA P35906 felis silve
11	30	73.2	407	1	RFC_SALTY P26479 salmonella
12	30	73.2	420	1	NAP1_HUMAN Q96009 homo sapien
13	30	73.2	436	1	NPT3_HUMAN Q00624 homo sapien
14	30	73.2	550	1	YM40_YEAST Q03212 saccharomyc
15	30	73.2	769	1	COMP_BACSU Q99027 bacillus su
16	30	73.2	775	1	THI1_SCHPO P36598 schizosacch
17	30	73.2	856	1	VPH1_NEUCR Q01290 neurospora
18	30	73.2	1388	1	CAIE_HUMAN P39059 homo sapien
19	29	70.7	80	1	YOT0_CAEEL P34655 caenorhabdi
20	29	70.7	216	1	TPIS_THEAC Q9hlb6 thermoplasm
21	29	70.7	325	1	MC5R_HUMAN P33032 homo sapien
22	29	70.7	325	1	MC5R_PANTR Q9tt23 pan troglod
23	29	70.7	487	1	MOT5_HUMAN O15374 homo sapien
24	29	70.7	602	1	CRK_DAUCA P53681 daucus caro
25	29	70.7	891	1	YB33_SCHPO O14338 schizosacch
26	29	70.7	923	1	YAUJA_SCHPO Q10165 schizosacch
27	29	70.7	4543	1	LRP1_CHICK P98157 gallus gall
28	28	68.3	79	1	PSPB_PIG P15782 sus scrofa
29	28	68.3	125	1	YN25_CAEEL P34584 caenorhabdi
30	28	68.3	146	1	HBB_MEGLY P11752 megaderma l
31	28	68.3	157	1	Y173_METJA Q57637 methanococc
32	28	68.3	171	1	IR10_HCMVA P16808 human cytom
33	28	68.3	178	1	LACB_PIG P04119 sus scrofa

ALIGNMENTS

RESULT 1					
CALH_MOUSE					
ID	CALH_MOUSE	STANDARD;	PRT;	1527 AA.	
AC	P39061; Q62002; Q61437;				
DT	01-FEB-1995 (Rel. 31, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].				
GN	COL18A1.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A. (SHORT ISOFORM).				
RC	STRAIN=BALB/C; TISSUE=Liver;				
RX	MEDLINE=94245707; PubMed=8188673;				
RA	Rehn M.V., Hintikka E., Pihlajaniemi T.;				
RT	"Primary structure of the alpha 1 chain of mouse type XVIII collagen, partial structure of the corresponding gene, and comparison of the alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen chain.";				
RL	J. Biol. Chem. 269:13929-13935(1994).				
RN	[2]				
RP	SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).				
RA	Rehn M., Hintikka E., Pihlajaniemi T.;				
RT	"Characterization of the mouse gene for the alpha-1 chain of type XVIII collagen (COL18A1) reveals that the three variant N-terminal polypeptide forms are transcribed from two widely separated promoters.";				
RL	Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).				
RX	MEDLINE=94240112; PubMed=8183894;				
RA	Rehn M.V., Pihlajaniemi T.;				
RT	"Alpha 1(XVIII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).				
RN	[4]				
RP	SEQUENCE OF 240-1527 FROM N.A.				
RC	TISSUE=Liver;				
RX	MEDLINE=94240111; PubMed=8183893;				
RA	Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;				
RT	"Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).				
RN	[5]				
RP	CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.				
RX	MEDLINE=97160848; PubMed=9008168;				
RA	O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S., Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;				
RT	"Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.";				
RL	Cell 88:277-285(1997).				

34	28	68.3	213	1	ATI2_VACCV	P24758 vaccinia vi
35	28	68.3	236	1	VHEL_WCMVM	P09499 white clove
36	28	68.3	236	1	VHEL_WCMVO	P15403 white clove
37	28	68.3	253	1	TPIS_BORBU	Q59182 borrelia bu
38	28	68.3	265	1	CTRC_NEIMB	P32015 neisseria m
39	28	68.3	269	1	IL1B_TRIVU	Q9xs77 trichosurus
40	28	68.3	294	1	VFEC_YERPE	Q56954 yersinia pe
41	28	68.3	305	1	CAG7_RAT	Q64686 rattus norv
42	28	68.3	322	1	ATI2_VACCC	P21114 vaccinia vi
43	28	68.3	325	1	MC5R_BOVIN	P56451 bos taurus
44	28	68.3	340	1	YDDR_ECOLI	P77308 escherichia
45	28	68.3	353	1	GPRO_RAT	P97639 rattus norv

[6]
X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
MEDLINE=98169382; PubMed=9501087;
Hohenester E., Sasaki T., Olsen B.R., Timpl R.;
"Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
resolution";
EMBO J. 17:1656-1664(1998).
-!- FUNCTION: ENDOSTATIN POTENTLY INHIBITS ENDOTHELIAL CELL
PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH
FACTOR SIGNALLING.
-!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM
(SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
-!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
-!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
INTERRUPTED HELICES (FACIT) FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L16898; AAA37434.1; -
EMBL; U03714; AAA20657.1; -
EMBL; U03715; AAC52901.1; -
EMBL; U34606; AAC52901.1; JOINED.
EMBL; U34608; AAC52901.1; JOINED.
EMBL; U34609; AAC52901.1; JOINED.
EMBL; U34610; AAC52901.1; JOINED.
EMBL; U34611; AAC52901.1; JOINED.
EMBL; U34612; AAC52901.1; JOINED.
EMBL; U34613; AAC52901.1; JOINED.
EMBL; U03716; AAC52901.1; JOINED.
EMBL; U03718; AAC52901.1; JOINED.
EMBL; U03715; AAC52902.1; -
EMBL; U34607; AAC52902.1; JOINED.
EMBL; U34608; AAC52902.1; JOINED.
EMBL; U34609; AAC52902.1; JOINED.
EMBL; U34610; AAC52902.1; JOINED.
EMBL; U34611; AAC52902.1; JOINED.
EMBL; U34612; AAC52902.1; JOINED.
EMBL; U34613; AAC52902.1; JOINED.
EMBL; U03716; AAC52902.1; JOINED.
EMBL; U03718; AAC52902.1; JOINED.
EMBL; U11636; AAC52178.1; -
EMBL; L22545; AAA19787.1; -
PDB; 1KOE; 16-FEB-99.
MGI; 88451; Col18a1.
InterPro; IPR000087; Collagen.
InterPro; IPR001791; Laminin_G.
InterPro; IPR003129; TSPN.
Pfam; PF01391; Collagen; 8.
Pfam; PF02210; TSPN; 2.
SMART; SM00282; LamG; 1.
SMART; SM00210; TSPN; 1.
Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
3D-structure.
SIGNAL 1 26 POTENTIAL.
CHAIN 27 1527 COLLAGEN ALPHA 1(XVIII) CHAIN.
CHAIN 1344 1527 ENDOSTATIN.
DOMAIN 27 538 NONHELICAL REGION 1 (NC1).
DOMAIN 539 565 TRIPLE-HELICAL REGION 1 (COL1).
DOMAIN 566 575 TRIPLE-HELICAL REGION 2 (NC2).
DOMAIN 576 649 TRIPLE-HELICAL REGION 2 (COL2).
DOMAIN 650 673 NONHELICAL REGION 3 (NC3).
DOMAIN 674 795 TRIPLE-HELICAL REGION 3 (COL3).
DOMAIN 796 818 NONHELICAL REGION 4 (NC4).
DOMAIN 819 901 TRIPLE-HELICAL REGION 4 (COL4).

FT DOMAIN 902 915 NONHELICAL REGION 5 (NC5).
FT DOMAIN 916 957 TRIPLE-HELICAL REGION 5 (COL5).
FT DOMAIN 958 970 NONHELICAL REGION 6 (NC6).
FT DOMAIN 971 1043 TRIPLE-HELICAL REGION 6 (COL6).
FT DOMAIN 1044 1053 NONHELICAL REGION 7 (NC7).
FT DOMAIN 1054 1086 TRIPLE-HELICAL REGION 7 (COL7).
FT DOMAIN 1087 1098 NONHELICAL REGION 8 (NC8).
FT DOMAIN 1099 1122 TRIPLE-HELICAL REGION 8 (COL8).
FT DOMAIN 1123 1129 NONHELICAL REGION 9 (NC9).
FT DOMAIN 1130 1181 TRIPLE-HELICAL REGION 9 (COL9).
FT DOMAIN 1182 1194 NONHELICAL REGION 10 (NC10).
FT DOMAIN 1195 1212 TRIPLE-HELICAL REGION 10 (COL10).
FT DOMAIN 1213 1527 NONHELICAL REGION 11 (NC11).
FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 700 700 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 1376 1516 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 1478 1508
FT SITE 1104 1106
FT VARSPLIC 1 212
FT VARSPLIC 213 238
FT
FT CONFLICT 900 900
FT CONFLICT 947 947
FT CONFLICT 964 964
FT CONFLICT 1157 1157
FT CONFLICT 1266 1266
FT CONFLICT 1276 1276
FT CONFLICT 1437 1437
SQ SEQUENCE 1527 AA; 156008 MW; 9645045AF140B513 CRC64;

Query Match 100.0%; Score 41; DB 1; Length 1527;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 1511 SYIVLCIE 1518

RESULT 2
CALH_HUMAN STANDARD; PRT; 1516 AA.
ID CALH_HUMAN Q9Y6Q8; Q9Y6Q7; Q9UK38;
AC P39060; Q9Y6Q8; Q9Y6Q7; Q9UK38;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN COL18A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98164096; PubMed=9503365;
RA Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;
RT "Complete primary structure of two variant forms of human type XVIII
collagen and tissue-specific differences in the expression of the
corresponding transcripts.";
RL Matrix Biol. 16:319-328(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Toudien S., Blechschmidt K., Polley A.,
RA Ohki M., Takagi T., Sakaki Y., Kumpf K., Lehmann R., Patterson D.,
RA Menzel U., Delabar J., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Reichwald K., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Rosenthal A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Shintani A., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Minoshima S., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Scharfe M.,

RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RN SEQUENCE OF 834-1516 FROM N.A.
RX MEDLINE=94245237; PubMed=8188291;
RA Oh S.P., Warman M.L., Seidin M.F., Cheng S., Knoll J.H., Timmons S.,
RA Olsen B.R.;
RT "Cloning of cDNA and genomic DNA encoding human type XVIII collagen
RT and localization of the alpha 1(XVIII) collagen gene to mouse
RT chromosome 10 and human chromosome 21.";
RL Genomics 19:494-499(1994).
RN [4]
RN SEQUENCE OF 1334-1516 FROM N.A.
RC TISSUE=Placenta;
RA Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
RT "Cloning and expression of human endostatin gene in Escherichia
RT coli.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RN INVOLVEMENT IN KNOBLOCH SYNDROME.
RX MEDLINE=20400145; PubMed=10942434;
RA Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
RA Passos-Bueno M.R.;
RT "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
RT and tumor growth, plays a critical role in the maintenance of retinal
RT structure and in neural tube closure.";
RL Hum. Mol. Genet. 9:2051-2058(2000).
RN [6]
RN VARIANT ASN-1437.
RX PubMed=11606364;
RA Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,
RA Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
RA Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RT "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
RT for the development of prostatic adenocarcinoma.";
RL Cancer Res. 61:7375-7378(2001).
CC -!- FUNCTION: COLA18A PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE
CC RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.
CC -!- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL
CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH
CC FACTOR SIGNALLING.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM/NC1-303 AND A LONG
CC FORM/NC-493 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS
CC IN LIVER, LUNG AND KIDNEY.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- POLYMORPHISM: There is an association between a polymorphism in
CC position 1437 and prostate cancer. Heterozygous Asn-1437
CC individuals have a 2.5 times increased chance of developing
CC prostate cancer as compared with homozygous Asp-1437 individuals.
CC -!- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KS
CC or KNO); an autosomal recessive disorder defined by the occurrence
CC of high myopia, vitreoretinal degeneration with retinal
CC detachment, macular abnormalities and occipital encephalocele.
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF018081; AAC39658.1; -
CC EMBL; AF018082; AAC39659.1; -
CC EMBL; AL163302; CAB90482.1; -

DR EMBL; L22548; AAF51864.1; -
DR EMBL; AF184060; AAF01310.1; ALT_INIT.
DR HSP; P39061; 1KOE.
DR GlycoSuiteDB; P39060; -
DR MIM; 120328; -
DR MIM; 267750; -
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 6.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW Polymorphism.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 1516 COLLAGEN ALPHA 1(XVIII) CHAIN.
FT CHAIN 1334 1516 ENDOSTATIN.
FT DOMAIN 24 516 NONHELICAL REGION 1 (NC1).
FT DOMAIN 517 550 TRIPLE-HELICAL REGION 1 (COL1).
FT DOMAIN 551 560 NONHELICAL REGION 2 (NC2).
FT DOMAIN 561 640 TRIPLE-HELICAL REGION 2 (COL2).
FT DOMAIN 641 664 NONHELICAL REGION 3 (NC3).
FT DOMAIN 665 786 TRIPLE-HELICAL REGION 3 (COL3).
FT DOMAIN 787 809 NONHELICAL REGION 4 (NC4).
FT DOMAIN 810 892 TRIPLE-HELICAL REGION 4 (COL4).
FT DOMAIN 893 906 NONHELICAL REGION 5 (NC5).
FT DOMAIN 907 948 TRIPLE-HELICAL REGION 5 (COL5).
FT DOMAIN 949 961 NONHELICAL REGION 6 (NC6).
FT DOMAIN 962 1034 TRIPLE-HELICAL REGION 6 (COL6).
FT DOMAIN 1035 1044 NONHELICAL REGION 7 (NC7).
FT DOMAIN 1045 1077 TRIPLE-HELICAL REGION 7 (COL7).
FT DOMAIN 1078 1089 NONHELICAL REGION 8 (NC8).
FT DOMAIN 1090 1111 TRIPLE-HELICAL REGION 8 (COL8).
FT DOMAIN 1112 1118 NONHELICAL REGION 9 (NC9).
FT DOMAIN 1119 1173 TRIPLE-HELICAL REGION 9 (COL9).
FT DOMAIN 1174 1186 NONHELICAL REGION 10 (NC10).
FT DOMAIN 1187 1204 TRIPLE-HELICAL REGION 10 (COL10).
FT DOMAIN 1205 1516 NONHELICAL REGION 11 (NC11).
FT CARBOHYD 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 691 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1329 O-LINKED (GALNAC. . .).
FT DISULFID 1366 /FTID=CAR_000150.
FT DISULFID 1468 BY SIMILARITY.
FT SITE 1095 BY SIMILARITY.
FT VARSPLIC 1 CELL ATTACHMENT SITE (POTENTIAL).
FT VARSPLIC 181 MISSING (IN SHORT ISOFORM).
FT VARIANT 1437 HTTEAGTLPAPTSPSLGRWPAPLTGPSVPPSS -> MA
FT PRCPWPWPRRRRLDLVAPLVLLGLVRAAEP (IN
FT SHORT ISOFORM).
FT D -> N.
FT /FTID=VAR_012709.
FT F -> S (IN REF. 2).
FT I -> V (IN REF. 2).
FT V -> L (IN REF. 3).
FT P -> R (IN REF. 3).
FT P -> R (IN REF. 3).
FT R -> L (IN REF. 3).
FT P -> L (IN REF. 3).
FT A -> P (IN REF. 3).
FT L -> K (IN REF. 3).
FT P -> A (IN REF. 3).
FT P -> A (IN REF. 3).
FT P -> PGP (IN REF. 2).
FT G -> GQ (IN REF. 3).
FT R -> G (IN REF. 3).
FT A -> G (IN REF. 3).
FT LR -> CG (IN REF. 3).
FT R -> T (IN REF. 4).
FT S -> Y (IN REF. 4).

SQ	SEQUENCE	1516 AA;	153840 MW;	3C70F29A4476EE76 CRC64;	
	Query Match	92.7%;	Score 38;	DB 1;	Length 1516;
	Best Local Similarity	87.5%;	Pred. No. 7.6;		
	Matches	7;	Conservative	1;	Mismatches 0; Indels 0; Gaps 0;
QY	1 SYIVLCIE 8				
	:				
Db	1501 AYIVLCIE 1508				
RESULT	3				
YAL7_HUMAN	STANDARD;	PRT;	1015 AA.		
AC	Q9UPZ3;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Hypothetical protein KIAA1017.				
GN	KIAA1017.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RX	MEDLINE=99246063; PubMed=10231032;				
RA	Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,				
RA	Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;				
RT	"Prediction of the coding sequences of unidentified human genes. XIII.				
RT	The complete sequences of 100 new cDNA clones from brain which code				
RT	for large proteins in vitro.";				
RL	DNA Res. 6:63-70(1999).				
CC	-!- SIMILARITY: TO HUMAN KIAA0297/KIAA0329.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AB023234; BAA76861.1; -				
KW	Hypothetical protein.				
SQ	SEQUENCE 1015 AA; 114842 MW; 0214C2BFD67EA426 CRC64;				
	Query Match	78.0%;	Score 32;	DB 1;	Length 1015;
	Best Local Similarity	57.1%;	Pred. No. 80;		
	Matches	4;	Conservative	3;	Mismatches 0; Indels 0; Gaps 0;
QY	2 YIVLCIE 8				
	: :				
Db	875 YLILCLE 881				
RESULT	4				
GLG1_CHICK	STANDARD;	PRT;	1142 AA.		
AC	Q02391; Q91019;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Golgi apparatus protein 1 precursor (Cysteine-rich fibroblast growth				
DE	factor receptor).				
GN	GLG1 OR CFR.				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC	Gallus.				

OX	NCBI_TaxID=9031;				
RN	[1]				
RP	SEQUENCE FROM N.A.; AND CHARACTERIZATION.				
RC	TISSUE=Embryonic brain;				
RX	MEDLINE=93078761; PubMed=1448090;				
RA	Burrus L.W., Zuber M.E., Lueddecke B.A., Olwin B.B.;				
RT	"Identification of a cysteine-rich receptor for fibroblast growth				
RT	factors.";				
RL	Mol. Cell. Biol. 12:5600-5609(1992).				
RN	[2]				
RP	SEQUENCE FROM N.A.; VARIANT, AND SUBCELLULAR LOCATION.				
RX	MEDLINE=97219477; PubMed=9066777;				
RA	Zuber M.E., Zhou Z., Burrus L.W., Olwin B.B.;				
RT	"Cysteine-rich FGF receptor regulates intracellular FGF-1 and FGF-2				
RT	levels.";				
RL	J. Cell. Physiol. 170:217-227(1997).				
RN	[3]				
RP	GLYCOSYLATION.				
RX	MEDLINE=90036970; PubMed=2553717;				
RA	Burrus L.W., Olwin B.B.;				
RT	"Isolation of a receptor for acidic and basic fibroblast growth factor				
RT	from embryonic chick.";				
RL	J. Biol. Chem. 264:18647-18653(1989).				
CC	-!- FUNCTION: BINDS E-SELECTIN (CELL-ADHESION LECTIN ON ENDOTHELIAL				
CC	CELLS MEDIATING THE BINDING OF NEUTROPHILS) (BY SIMILARITY). BINDS				
CC	FIBROBLAST GROWTH FACTOR (FGF). MAY BE INVOLVED IN INTRACELLULAR				
CC	FGF TRAFFICKING AND THE REGULATION OF CELLULAR RESPONSES TO FGFS.				
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein; Golgi.				
CC	-!- PTM: Fucosylation is essential for binding to E-Selectin (By				
CC	similarity).				
CC	-!- PTM: N-glycosylated. Contains sialic acid residues.				
CC	-!- MISCELLANEOUS: THE CHO.MUTCFR MUTANT IS UNABLE TO REGULATE FGF				
CC	LEVELS AND IS DETECTED THROUGHOUT THE CELL.				
CC	-!- SIMILARITY: CONTAINS 15 CYSTEINE-RICH GLG1 REPEATS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M95766; AAA48769.1; -				
DR	EMBL; U48395; AAB39211.1; -				
DR	InterPro; IPR001893; Cys_rich_FGFR.				
DR	Pfam; PF00839; cys_rich_FGFR; 15.				
KW	Signal; Transmembrane; Sialic acid; Glycoprotein; Golgi stack; Repeat.				
FT	SIGNAL 1 29				POTENTIAL.
FT	CHAIN 30 1142				GOLGI APPARATUS PROTEIN 1.
FT	DOMAIN 30 1108				EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 1109 1129				POTENTIAL.
FT	DOMAIN 1130 1142				CYTOPLASMIC (POTENTIAL).
FT	REPEAT 116 175				CYS-RICH GLG1 1.
FT	REPEAT 177 241				CYS-RICH GLG1 2.
FT	REPEAT 252 309				CYS-RICH GLG1 3.
FT	REPEAT 312 376				CYS-RICH GLG1 4.
FT	REPEAT 380 436				CYS-RICH GLG1 5.
FT	REPEAT 441 500				CYS-RICH GLG1 6.
FT	REPEAT 503 567				CYS-RICH GLG1 7.
FT	REPEAT 575 631				CYS-RICH GLG1 8.
FT	REPEAT 632 692				CYS-RICH GLG1 9.
FT	REPEAT 695 751				CYS-RICH GLG1 10.
FT	REPEAT 762 819				CYS-RICH GLG1 11.
FT	REPEAT 820 875				CYS-RICH GLG1 12.
FT	REPEAT 878 942				CYS-RICH GLG1 13.
FT	REPEAT 945 1005				CYS-RICH GLG1 14.
FT	REPEAT 1007 1064				CYS-RICH GLG1 15.
FT	DOMAIN 64 72				POLY-GLY.
FT	CARBOHYD 128 128				N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 173 173				N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 544 544				N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 640 640				N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 749 749 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 1042 1142 CAAIPGGRGMSCLMEALDKRVLQPECKKRLNDRIEMW
FT SYAAKVAPAEGFSDLAMQVMTSPSKNYILSVITVGLCVLFL
FT IGLMCGRTKRVTRDKDR -> IVLKCGAMLQRLPQKAS
FT LTLPCCKLRPLRPRTYCL (IN CHO.MUTCFR
FT MUTANT).
SQ SEQUENCE 1142 AA; 129709 MW; 91CC9BE8D53CABB8 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 1142;
Best Local Similarity 50.0%; Pred. No. 89;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 362 SYLLMCLE 369
||::||:|

RESULT 5
GLGL_CRIGR STANDARD; PRT; 1160 AA.
AC Q921E9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Golgi apparatus protein 1 precursor (Golgi sialoglycoprotein MG-160)
DE (E-selectin ligand 1) (ESL-1) (Latent TGF-Beta complexed protein-1)
DE (LTCP-1).
GN GLG1 OR MG160 OR ESL1.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 80-87; 180-201; 215-223; 240-252;
RP 266-273; 342-348; 442-448; 527-536; 632-645; 847-858; 930-937 AND
RP 961-977.
RC TISSUE=Ovary;
RX MEDLINE=97307852; PubMed=9182700;
RA Olofsson A., Hellman U., Ten Dijke P., Grimsby S., Ichijo H.,
RA Moren A., Miyazono K., Heldin C.-H.;
RT "Latent transforming growth factor-beta complex in Chinese hamster
RT ovary cells contains the multifunctional cysteine-rich fibroblast
RT growth factor receptor, also termed E-selectin-ligand or MG-160.";
RL Biochem. J. 324:427-434(1997).
CC -!- FUNCTION: BINDS FIBROBLAST GROWTH FACTOR AND E-SELECTIN (CELL-
CC ADHESION LECTIN ON ENDOTHELIAL CELLS MEDIATING THE BINDING OF
CC NEUTROPHILS) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein; Golgi (By
CC similarity).
CC -!- PTM: Fucosylation is essential for binding to E-Selectin (By
CC similarity).
CC -!- PTM: N-glycosylated. Contains sialic acid residues (By
CC similarity).
CC -!- SIMILARITY: CONTAINS 15 CYSTEINE-RICH GLG1 REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U51162; AAD00079.1; .
DR InterPro; IPR001893; Cys_rich_FGFR.
DR Pfam; PF00839; cys_rich_FGFR; 15.
KW Signal; Transmembrane; Sialic acid; Glycoprotein; Golgi stack; Repeat.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 1160 GOLGI APPARATUS PROTEIN 1.
FT DOMAIN 19 1126 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1127 1147 POTENTIAL.

FT DOMAIN 1148 1160 CYTOPLASMIC (POTENTIAL).
FT REPEAT 134 193 CYS-RICH GLG1 1.
FT REPEAT 195 259 CYS-RICH GLG1 2.
FT REPEAT 270 327 CYS-RICH GLG1 3.
FT REPEAT 330 394 CYS-RICH GLG1 4.
FT REPEAT 398 454 CYS-RICH GLG1 5.
FT REPEAT 459 518 CYS-RICH GLG1 6.
FT REPEAT 521 585 CYS-RICH GLG1 7.
FT REPEAT 593 649 CYS-RICH GLG1 8.
FT REPEAT 650 710 CYS-RICH GLG1 9.
FT REPEAT 713 769 CYS-RICH GLG1 10.
FT REPEAT 780 837 CYS-RICH GLG1 11.
FT REPEAT 838 893 CYS-RICH GLG1 12.
FT REPEAT 896 960 CYS-RICH GLG1 13.
FT REPEAT 963 1023 CYS-RICH GLG1 14.
FT REPEAT 1025 1082 CYS-RICH GLG1 15.
FT DOMAIN 61 67 POLY-GLN.
FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 562 562 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 767 767 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1160 AA; 132325 MW; 9D30E2D1DAAC03CB CRC64;

Query Match 78.0%; Score 32; DB 1; Length 1160;
Best Local Similarity 50.0%; Pred. No. 91;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 380 SYLLMCLE 387
||::||:|

RESULT 6
GLGL_RAT STANDARD; PRT; 1171 AA.
ID GLGL_RAT
AC Q62638;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Golgi apparatus protein 1 precursor (Golgi sialoglycoprotein MG-160)
DE (E-selectin ligand 1) (ESL-1).
GN GLG1 OR MG160 OR ESL1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 90-103; 174-189; 225-241; 494-508;
RP 538-547; 687-700; 993-1010 AND 1104-1115.
RC STRAIN=Sprague-Dawley; TISSUE=Brain cortex;
RX MEDLINE=95286693; PubMed=7768993;
RA Gonatas J.O., Mourelatos Z., Stieber A., Lane W.S., Brosius J.,
RA Gonatas N.K.;
RT "MG-160, a membrane sialoglycoprotein of the medial cisternae of the
RT rat Golgi apparatus, binds basic fibroblast growth factor and exhibits
RT a high level of sequence identity to a chicken fibroblast growth
RT factor receptor.";
RL J. Cell Sci. 108:457-467(1995).
RN [2]
RN CHARACTERIZATION, AND SUBCELLULAR LOCATION.
RX MEDLINE=89079727; PubMed=2909545;
RA Gonatas J.O., Mezitis S.G.E., Stieber A., Fleischer B., Gonatas N.K.;
RT "MG-160. A novel sialoglycoprotein of the medial cisternae of the
RT Golgi apparatus.";
RL J. Biol. Chem. 264:646-653(1989).
CC -!- FUNCTION: BINDS FIBROBLAST GROWTH FACTOR. BINDS E-SELECTIN (CELL-
CC ADHESION LECTIN ON ENDOTHELIAL CELLS MEDIATING THE BINDING OF
CC NEUTROPHILS) (By similarity).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN; GOLGI MEDIAL
CC CISTERNAE.
CC -!- TISSUE SPECIFICITY: Widely expressed; found in kidney, pancreatic

CC islets, parathyroid, thyroid, adrenal tissue, brain neurons,
CC astrocytes, adenohypophysis, cultured pheochromocytoma cells.
CC -!- PTM: Fucosylation is essential for binding to E-Selectin (By
CC similarity).
CC -!- PTM: N-glycosylated. Contains sialic acid residues.
CC -!- SIMILARITY: CONTAINS 15 CYSTEINE-RICH GLG1 REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U08136; AAB03365.1; -.
DR InterPro: IPR001893; Cys_rich_FGFR.
DR Pfam; PF00839; cys_rich_FGFR; 15.
KW Signal; Transmembrane; Sialic acid; Glycoprotein; Golgi stack; Repeat.
FT SIGNAL 1 32
FT CHAIN 33 1171
FT DOMAIN 33 1137 GOLGI APPARATUS PROTEIN 1.
FT TRANSMEM 1138 1158 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1159 1171 POTENTIAL.
FT REPEAT 145 204 CYTOPLASMIC (POTENTIAL).
FT REPEAT 206 270 CYS-RICH GLG1 1.
FT REPEAT 281 338 CYS-RICH GLG1 2.
FT REPEAT 341 405 CYS-RICH GLG1 3.
FT REPEAT 409 465 CYS-RICH GLG1 4.
FT REPEAT 470 529 CYS-RICH GLG1 5.
FT REPEAT 532 596 CYS-RICH GLG1 6.
FT REPEAT 604 660 CYS-RICH GLG1 7.
FT REPEAT 661 721 CYS-RICH GLG1 8.
FT REPEAT 724 780 CYS-RICH GLG1 9.
FT REPEAT 791 848 CYS-RICH GLG1 10.
FT REPEAT 849 904 CYS-RICH GLG1 11.
FT REPEAT 907 971 CYS-RICH GLG1 12.
FT REPEAT 974 1034 CYS-RICH GLG1 13.
FT REPEAT 1036 1093 CYS-RICH GLG1 14.
FT DOMAIN 66 72 CYS-RICH GLG1 15.
FT DOMAIN 76 79 POLY-GLN.
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1171 AA; 133556 MW; CDD9D34D109E272F CRC64;

Query Match 78.0%; Score 32; DB 1; Length 1171;
Best Local Similarity 50.0%; Pred. No. 91;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 391 SYLLMCLE 398

RESULT 7
GLG1_MOUSE
ID GLG1_MOUSE STANDARD; PRT; 1175 AA.
AC Q61543; Q9QZ40;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Golgi apparatus protein 1 precursor (Golgi sialoglycoprotein MG-160)
DE (E-selectin ligand 1) (PSL-1) (Selel).
GN GLG1 OR MG160 OR SELEL OR ESL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 255-266; 357-363; 636-641 AND
RP 744-750.
RC TISSUE=Neutrophils;
RX MEDLINE=95157635; PubMed=7531823;
RA Steegmaier M., Levinovitz A., Isenmann S., Borges E., Lenter M.,
RA Kocher H.P., Kleuser B., Vestweber D.;
RT "The E-selectin-ligand ESL-1 is a variant of a receptor for fibroblast
RT growth factor.";
RL Nature 373:615-620(1995).
RN [2]
RP SEQUENCE OF 1-138 FROM N.A.
RC STRAIN=129/Sv; TISSUE=Embryonic stem cells;
RX MEDLINE=20028322; PubMed=10556428;
RA Willmroth F., Beaudet A.L.;
RT "Structure of the murine E-selectin ligand 1 (ESL-1) gene and
RT assignment to chromosome 8.";
RL Mamm. Genome 10:1085-1088(1999).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=97254533; PubMed=9099943;
RA Steegmaier M., Borges E., Berger J., Schwarz H., Vestweber D.;
RT "The E-selectin-ligand ESL-1 is located in the Golgi as well as on
RT microvilli on the cell surface.";
RL J. Cell Sci. 110:687-694(1997).
CC -!- FUNCTION: BINDS FIBROBLAST GROWTH FACTOR (BY SIMILARITY). BINDS
CC E-SELECTIN (CELL-ADHESION LECTIN ON ENDOTHELIAL CELLS MEDIATING
CC THE BINDING OF NEUTROPHILS).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN; GOLGI AND
CC MICROVILLI ON THE CELL SURFACE.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED; FOUND IN MYELOID CELLS,
CC FIBROBLASTS, COLON CARCINOMA, ENDOTHELIOOMA, TERATOCARCINOMA,
CC LYMPHOMA, MYELOMA.
CC -!- PTM: Fucosylation is essential for binding to E-Selectin.
CC -!- PTM: N-glycosylated. Contains sialic acid residues (By
CC similarity).
CC -!- SIMILARITY: CONTAINS 15 CYSTEINE-RICH GLG1 REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X84037; CAA58855.1; -.
DR EMBL; Y12462; CAA73066.1; -.
DR MGD; MGI:104967; Selel.
DR InterPro: IPR001893; Cys_rich_FGFR.
DR Pfam; PF00839; cys_rich_FGFR; 15.
KW Signal; Transmembrane; Sialic acid; Glycoprotein; Golgi stack; Repeat.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 1175 GOLGI APPARATUS PROTEIN 1.
FT DOMAIN 28 1141 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1142 1162 POTENTIAL.
FT DOMAIN 1163 1175 CYTOPLASMIC (POTENTIAL).
FT REPEAT 149 208 CYS-RICH GLG1 1.
FT REPEAT 210 274 CYS-RICH GLG1 2.
FT REPEAT 285 342 CYS-RICH GLG1 3.
FT REPEAT 345 409 CYS-RICH GLG1 4.
FT REPEAT 413 469 CYS-RICH GLG1 5.
FT REPEAT 474 533 CYS-RICH GLG1 6.
FT REPEAT 536 600 CYS-RICH GLG1 7.
FT REPEAT 608 664 CYS-RICH GLG1 8.
FT REPEAT 665 725 CYS-RICH GLG1 9.
FT REPEAT 728 784 CYS-RICH GLG1 10.
FT REPEAT 795 852 CYS-RICH GLG1 11.
FT REPEAT 853 908 CYS-RICH GLG1 12.
FT REPEAT 911 975 CYS-RICH GLG1 13.
FT REPEAT 978 1038 CYS-RICH GLG1 14.
FT REPEAT 1040 1097 CYS-RICH GLG1 15.
FT DOMAIN 46 51 POLY-GLY.
FT DOMAIN 66 70 POLY-GLN.

FT DOMAIN 74 82 POLY-GLN.
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 577 577 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 673 673 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1175 AA; 133733 MW; 1058355DD38C7338B CRC64;

Query Match 78.0%; Score 32; DB 1; Length 1175;
Best Local Similarity 50.0%; Pred. No. 92;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 395 SYLLMCLE 402

RESULT 8
GLG1_HUMAN STANDARD; PRT; 1179 AA.
AC Q92896; Q13221;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Golgi apparatus protein 1 precursor (Golgi sialoglycoprotein MG-160)
DE (E-selectin ligand 1) (ESL-1) (Cysteine-rich fibroblast growth factor receptor) (CFR-1).
DE GLG1 OR MGL60 OR ESL1 OR CFR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphoblast, and Fetal brain;
RX MEDLINE=97138100; PubMed=8985126;
RA Mourelatos Z., Gonatas J.O., Cinato E., Gonatas N.K.;
RT "Cloning and sequence analysis of the human MGL60, a fibroblast growth factor and E-selectin binding membrane sialoglycoprotein of the Golgi apparatus."
RL DNA Cell Biol. 15:1121-1128(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Wu M., Chen J., Tan Y.H., Hong W.J., Ting R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=90285470; PubMed=2355176;
RA Croud S., Mezitis S.G.E., Stieber A., Chen Y.J., Gonatas J.O.,
RA Goud B., Gonatas N.K.;
RT "Immunocytochemical visualization of the Golgi apparatus in several species, including human, and tissues with an antiserum against MG-160, a sialoglycoprotein of rat Golgi apparatus."
RL J. Histochem. Cytochem. 38:957-963(1990).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=97307852; PubMed=9182700;
RA Olofsson A., Hellman U., Ten Dijke P., Grimsby S., Ichijo H.,
RA Moren A., Miyazono K., Heldin C.H.;
RT "Latent transforming growth factor-beta complex in Chinese hamster ovary cells contains the multifunctional cysteine-rich fibroblast growth factor receptor, also termed E-selectin-ligand or MG-160.";
RL Biochem. J. 324:427-434(1997).
CC -!- FUNCTION: BINDS FIBROBLAST GROWTH FACTOR AND E-SELECTIN (CELL-ADHESION LECTIN ON ENDOTHELIAL CELLS MEDIATING THE BINDING OF NEUTROPHILS).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN; GOLGI.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN PANCREAS, SKELETAL MUSCLE, PLACENTA, HEART, TESTIS AND OVARY. ALSO FOUND IN THE KIDNEY, LIVER, LUNG AND BRAIN.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED BOTH IN ADULT AND FETAL TISSUES.
CC -!- PTM: Fucosylation is essential for binding to E-selectin (By

CC similarity).
CC -!- PTM: N-glycosylated. Contains sialic acid residues (By similarity).
CC -!- SIMILARITY: CONTAINS 15 CYSTEINE-RICH GLG1 REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; U64791; AAB06460.1; -;
CC EMBL; U28811; AAB02178.1; -;
CC MIM; 600753; -;
CC InterPro; IPR001893; Cys_rich_FGFR.
CC Pfam; PF00839; cys_rich_FGFR; 15.
KW Signal; Transmembrane; Sialic acid; Glycoprotein; Golgi stack; Repeat.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 1179 GOLGI APPARATUS PROTEIN 1.
FT DOMAIN 30 1145 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1146 1166 POTENTIAL.
FT DOMAIN 1167 1179 CYTOPLASMIC (POTENTIAL).
FT REPEAT 153 212 CYS-RICH GLG1 1.
FT REPEAT 214 278 CYS-RICH GLG1 2.
FT REPEAT 289 346 CYS-RICH GLG1 3.
FT REPEAT 349 413 CYS-RICH GLG1 4.
FT REPEAT 417 473 CYS-RICH GLG1 5.
FT REPEAT 478 537 CYS-RICH GLG1 6.
FT REPEAT 540 604 CYS-RICH GLG1 7.
FT REPEAT 612 668 CYS-RICH GLG1 8.
FT REPEAT 669 729 CYS-RICH GLG1 9.
FT REPEAT 732 788 CYS-RICH GLG1 10.
FT REPEAT 799 856 CYS-RICH GLG1 11.
FT REPEAT 857 912 CYS-RICH GLG1 12.
FT REPEAT 915 979 CYS-RICH GLG1 13.
FT REPEAT 982 1042 CYS-RICH GLG1 14.
FT REPEAT 1044 1101 CYS-RICH GLG1 15.
FT DOMAIN 54 57 POLY-GLY.
FT DOMAIN 71 86 POLY-GLN.
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 581 581 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 786 786 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 27 35 AEKLPGHV -> GRNSPARAS (IN REF. 2).
FT CONFLICT 67 67 L -> P (IN REF. 2).
FT CONFLICT 85 85 L -> P (IN REF. 2).
FT CONFLICT 99 99 MISSING (IN REF. 2).
FT CONFLICT 526 526 L -> S (IN REF. 2).
FT CONFLICT 702 702 N -> T (IN REF. 2).
FT CONFLICT 707 707 MISSING (IN REF. 2).
FT CONFLICT 1082 1082 I -> L (IN REF. 2).
SQ SEQUENCE 1179 AA; 134592 MW; F90061ACF6B42CF9 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 1179;
Best Local Similarity 50.0%; Pred. No. 92;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 399 SYLLMCLE 406

RESULT 9
YA36_METJA STANDARD; PRT; 233 AA.
ID YA36_METJA
AC Q58442;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```
DE Hypothetical protein MJ1036.
GN MJ1036.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=868087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC !- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67546; AAB99045.1; -.
CC TIGR; MJ1036; -.
CC InterPro; IPR003689; Zip.
CC Pfam; PF02535; Zip; 1.
CC KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 36 56 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
SQ SEQUENCE 233 AA; 26170 MW; D2D82ACF6200DB58 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 233;
Best Local Similarity 71.4%; Pred. No. 52;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
DB 15 SFIVMCI 21

RESULT 10
RDS_FELCA
ID RDS_FELCA STANDARD; PRT; 346 AA.
AC P35906;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peripherin (Retinal degeneration slow protein).
GN RDS.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ABYSSINIAN;
RX MEDLINE=94162777; PubMed=8118105;
RA Gorin M.B., Snyder S., To A.C., Narfstrom K., Curtis R.;
RT "The cat RDS transcript: candidate gene analysis and phylogenetic
```

```
RT sequence analysis."
RL Mamm. Genome 4:544-548(1993).
CC !- FUNCTION: MAY FUNCTION AS AN ADHESION MOLECULE INVOLVED IN
CC STABILIZATION AND COMPACTION OF OUTER SEGMENT DISKS OR IN THE
CC MAINTENANCE OF THE CURVATURE OF THE RIM. IT IS ESSENTIAL FOR DISK
CC MORPHOGENESIS.
CC !- SUBUNIT: HOMODIMER; DISULFIDE-LINKED. PROBABLY FORMS A COMPLEX
CC WITH A ROM1 HOMODIMER. OTHER PROTEINS COULD ASSOCIATE WITH THIS
CC COMPLEX IN RODS (BY SIMILARITY).
CC !- SUBCELLULAR LOCATION: Integral membrane protein.
CC !- TISSUE SPECIFICITY: RETINA (PHOTORECEPTOR). IN RIM REGION OF ROS
CC (ROD OUTER SEGMENT) DISKS.
CC !- SIMILARITY: BELONGS TO THE RDS(PERIPHERIN) / ROM1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M94047; AAA19175.1; -.
CC InterPro; IPR000830; RDS_ROM.
CC InterPro; IPR000301; Transmem_4.
CC Pfam; PF00335; transmembrane4; 1.
CC PRINTS; PR00218; PERIPHERNRDS.
CC PROSITE; PS00930; RDS_ROM1; 1.
CC KW Photoreceptor; Vision; Transmembrane; Glycoprotein.
FT DOMAIN 1 24 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 25 43 POTENTIAL.
FT DOMAIN 44 61 LUMENAL (POTENTIAL).
FT TRANSMEM 62 80 POTENTIAL.
FT DOMAIN 81 99 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 100 123 POTENTIAL.
FT DOMAIN 124 264 LUMENAL (POTENTIAL).
FT TRANSMEM 265 290 POTENTIAL.
FT DOMAIN 291 346 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 229 229 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 346 AA; 39171 MW; E700F0E29E4759A6 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 346;
Best Local Similarity 57.1%; Pred. No. 75;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
DB 100 SYLVVCV 106

RESULT 11
RFS_SALTY
ID RFS_SALTY STANDARD; PRT; 407 AA.
AC P26479;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE O-antigen polymerase.
GN RFS OR STM1332 OR STY1779.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium;
RX MEDLINE=91193209; PubMed=1707412;
RA Collins L.V., Hackett J.;
RT "Molecular cloning, characterization, and nucleotide sequence of the
RT rfc gene, which encodes an O-antigen polymerase of Salmonella
```

typhimurium.";
RL J. Bacteriol. 173:2521-2529(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSCI412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -!- FUNCTION: MAY LINK THE O-ANTIGEN TETRASACCHARIDE UNITS INTO LONG
CC CHAINS, GIVING RISE TO TYPICAL SMOOTH LPS.
CC -!- PATHWAY: LIPOPOLYSACCHARIDE BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC CONTAINS SEVERAL POTENTIAL TRANSMEMBRANE DOMAINS (POTENTIAL).
CC -!- SIMILARITY: TO OTHER O-ANTIGEN POLYMERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60066; AAA27210.1; -.
DR EMBL; AE008758; AAL20257.1; -.
DR EMBL; AL627271; CAD02021.1; -.
DR PIR; A43672; A43672.
DR StyGene; SG10352; rfc.
KW Lipopolysaccharide biosynthesis; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 2 22 POTENTIAL.
FT TRANSMEM 31 51 POTENTIAL.
FT TRANSMEM 63 83 POTENTIAL.
FT TRANSMEM 101 121 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
FT TRANSMEM 179 199 POTENTIAL.
FT TRANSMEM 211 231 POTENTIAL.
FT TRANSMEM 320 340 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.
FT TRANSMEM 382 402 POTENTIAL.
SQ SEQUENCE 407 AA; 47461 MW; AF187D8633D9CEDE CRC64;

Query Match 73.2%; Score 30; DB 1; Length 407;
Best Local Similarity 71.4%; Pred. No. 87;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIVLCI 7

||| ||:

Db 5 SYIALCL 11

RESULT 12
NAP1_HUMAN
ID NAP1_HUMAN STANDARD; PRT; 420 AA.
AC O96009;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Napsin 1 precursor (EC 3.4.23.-) (Napsin A) (NAPA) (TA01/TA02)
DE (Aspartyl protease 4) (Asp 4) (ASP4).
GN NAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung, and Kidney;
RX MEDLINE=99092989; PubMed=9877162;
RA Tatnell P.J., Powell D.J., Hill J., Smith T.S., Tew D.G., Kay J.;
RT "Napsins: new human aspartic proteinases. Distinction between two
RT closely related genes.";
RL FEBS Lett. 441:43-48(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Koelsch G., Wu S., Henthorn J., Tang J., Lin X.;
RT "New human aspartic proteases napsin 1 and napsin 2: molecular
RT cloning and intracellular localization of napsin 1.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=Fetal lung;
RX MEDLINE=20047840; PubMed=10580105;
RA Chuman Y., Bergman A.-C., Ueno T., Saito S., Sakaguchi K.,
RA Alaiya A.A., Franzen B., Bergman T., Arnott D., Auer G., Appella E.,
RA Joernvall H., Linder S.;
RT "Napsin A, a member of the aspartic protease family, is abundantly
RT expressed in normal lung and kidney tissue and is expressed in lung
RT adenocarcinomas.";
RL FEBS Lett. 462:129-134(1999).
RN [4]
RP SEQUENCE FROM N.A.
RA Bienkowski M.J., Shuck M.E., Slightom J.L., Drong R.F.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY BE INVOLVED IN PROCESSING OF PNEUMOCYTE SURFACTANT
CC PRECURSORS.
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN ADULT LUNG (TYPE II
CC PNEUMOCYTES) AND KIDNEY AND IN FETAL LUNG. LOW LEVELS IN ADULT
CC SPLEEN AND VERY LOW LEVELS IN PERIPHERAL BLOOD LEUKOCYTES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF090386; AAD04917.1; -.
DR EMBL; AF098484; AAD13215.1; -.
DR EMBL; AF200345; AAF17081.1; -.
DR HSSP; P00797; 2REN.
DR MEROPS; A01.046; -.
DR MIM; 605631; -.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 25 POTENTIAL.

```
FT PROPEP 26 63 ACTIVATION PEPTIDE.
FT CHAIN 64 420 NAPSIN 1.
FT ACT_SITE 96 BY SIMILARITY.
FT ACT_SITE 283 BY SIMILARITY.
FT DISULFID 109 116 BY SIMILARITY.
FT DISULFID 274 278 BY SIMILARITY.
FT DISULFID 317 354 BY SIMILARITY.
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 420 AA; 45386 MW; 018B86AE5BD0C865 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 420;
Best Local Similarity 71.4%; Pred. No. 89;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YIVLCIE 8
Db 313 YIILCE 319

RESULT 13
NPT3_HUMAN STANDARD; PRT; 436 AA.
AC O00624;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sodium-dependent phosphate transport protein 3 (Sodium/phosphate
DE cotransporter 3) (Na(+)/PI cotransporter 3).
GN SLC17A3 OR NPT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed-9149941;
RA Ruddy D.A., Kronmal G.S., Lee V.K., Mintier G.A., Quintana L.,
RA Domingo R. Jr., Meyer N.C., Irrinki A., McClelland E.E., Fulian A.,
RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchihashi Z.,
RA Wolff R.K., Schatzman R.C., Feder J.N.;
RT "A 1.1-Mb transcript map of the hereditary hemochromatosis locus.";
RL Genome Res. 7:441-456(1997).
CC -!- FUNCTION: MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE
CC INTO CELLS VIA NA+ COTransport (By SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
SQ SEQUENCE 436 AA; 47350 MW; DF02F618E83A572F CRC64;

Query Match 73.2%; Score 30; DB 1; Length 436;
Best Local Similarity 57.1%; Pred. No. 92;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
Db 193 SFILCV 199

RESULT 14
YM40_YEAST STANDARD; PRT; 550 AA.
ID YM40_YEAST Q03830;
AC Q03212; Q03830;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 62.5 kDa protein in ALD2-DDR48 intergenic region.
GN YMR171C OR YM8010.01C OR YM8520.20C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-241 FROM N.A.
RC STRAIN=S288C / AB972;
RA Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 215-550 FROM N.A.
RC STRAIN=S288C / AB972;
RA Hunt S., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO YEAST YKL124W AND S.POMBE SPCC285.10C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

DE Sensor protein comp (EC 2.7.1.1.-).
GN COMP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BD630;
RX MEDLINE=90337321; PubMed=2116363;
RA Weinrauch Y., Penchev R., Dubnau E., Smith I., Dubnau D.;
RT "A Bacillus subtilis regulatory gene product for genetic competence
RT and sporulation resembles sensor protein members of the bacterial
RT two-component signal-transduction systems.";
RL Genes Dev. 4:860-872(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Oudega B., Koningssteijn G., Van Zoest A.;
RT "Bacillus subtilis genome project, DNA sequence from yufA to
RT yufD.";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-16 FROM N.A.
RC STRAIN=168;
RX MEDLINE=91358358; PubMed=1715859;
RA Weinrauch Y., Msadek T., Kunst F., Dubnau D.;
RT "Sequence and properties of comQ, a new competence regulatory gene of
RT Bacillus subtilis.";
RL J. Bacteriol. 173:5685-5693(1991).
CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM COMP/COMA.
CC HISTIDINE KINASE THAT IS REQUIRED EARLY IN THE COMPETENCE CASCADE.
CC ACTIVATES COMA PROTEIN BY PHOSPHORYLATION. IT PLAYS ROLE IN
CC SPOULATION, AT LEAST PARTLY INTERCHANGEABLE WITH THAT OF SPOIIJ.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54010; CA37957.1; -.
DR EMBL; Z93932; CAB07903.1; -.
DR EMBL; M22856; AAA22319.1; -.
DR EMBL; M71283; AAA22324.1; -.
DR EMBL; Z99120; CAB15157.1; -.
DR PIR; A35848; A35848.
DR Subtilist; BG10380; comp.
DR InterPro; IPR003594; HATPase_c.
DR InterPro; IPR004359; HIS_KIN_sig.
DR Pfam; PF02518; HATPase_c; 1.
DR SMART; SM00387; HATPase_c; 1.
DR PROSITE; PS50109; HIS_KIN; 1.
KW Sensory transduction; Phosphorylation; Transferase; Kinase;
KW Transmembrane; Complete proteome.
FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 10 33 POTENTIAL.
FT DOMAIN 34 113 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 114 134 POTENTIAL.
FT DOMAIN 135 144 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 145 167 POTENTIAL.
FT DOMAIN 168 235 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 236 257 POTENTIAL.
FT DOMAIN 258 272 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 273 295 POTENTIAL.
FT DOMAIN 296 299 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 300 323 POTENTIAL.
FT DOMAIN 324 337 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 338 357 POTENTIAL.

FT DOMAIN 358 361 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 362 383 POTENTIAL.
FT DOMAIN 384 769 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 571 769 HISTIDINE KINASE.
FT MOD_RES 456 456 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 604 604 S -> C (IN REF. 1).
FT CONFLICT 610 610 D -> Y (IN REF. 1).
FT CONFLICT 628 628 E -> G (IN REF. 1).
FT CONFLICT 636 637 QL -> PV (IN REF. 1).
SQ SEQUENCE 769 AA; 89316 MW; ESCEACCE278DE5B4 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 769;
Best Local Similarity 57.1%; Pred. No. 1.5e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
II::II:
Db 497 SYVLLCL 503

Search completed: May 31, 2002, 10:33:26
Job time: 332 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:25:38 ; Search time 26.62 Seconds
(without alignments)
28.877 Million cell updates/sec

Title: US-09-589-777C-24
Perfect score: 41
Sequence: 1 SYIVLCIE 8
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	41	100.0	1315	2 A56101	collagen alpha 1(X
2	41	100.0	1774	2 B56101	collagen alpha 1(X
3	38	92.7	684	2 A53019	collagen alpha 1(X
4	32	78.0	147	2 F71498	hypothetical prote
5	32	78.0	147	2 E81656	conserved hypothet
6	32	78.0	293	2 D96605	unknown protein F1
7	32	78.0	1142	2 A45031	cysteine-rich fibr
8	32	78.0	1175	2 S52417	E-selectin ligand-
9	31	75.6	39	2 G85837	hypothetical prote
10	31	75.6	158	2 H72029	conserved hypothet
11	31	75.6	158	2 D86593	CT565 hypothetical
12	31	75.6	163	2 D84853	hypothetical prote
13	31	75.6	174	2 C97148	hypothetical prote
14	31	75.6	201	2 E90114	hypothetical prote
15	31	75.6	208	2 JC7755	bombinakinin M pre
16	31	75.6	235	2 T23501	hypothetical prote
17	31	75.6	316	2 B75323	conserved hypothet
18	31	75.6	466	2 B82482	alpha-amylase VCA0
19	31	75.6	479	2 F86285	F9L1.11 protein -
20	31	75.6	566	2 E81404	acetolactate synth
21	31	75.6	2342	2 T18200	fatty-acid synthas
22	30	73.2	96	2 T42273	hypothetical prote
23	30	73.2	129	2 T21687	hypothetical prote
24	30	73.2	141	2 C71605	clathrin coat asse
25	30	73.2	151	2 T48823	hypothetical prote
26	30	73.2	163	2 T33130	hypothetical prote
27	30	73.2	168	2 T12834	hypothetical prote
28	30	73.2	233	2 C64429	hypothetical prote
29	30	73.2	267	2 D83313	probable permease

30 30 73.2 277 2 D84596 hypothetical prote
31 30 73.2 346 2 I46087 peripherin - cat
32 30 73.2 353 2 G71567 probable cytochrom
33 30 73.2 407 2 A43672 O-antigen polymera
34 30 73.2 407 2 AD0706 O-antigen polymera
35 30 73.2 507 2 D86207 hypothetical prote
36 30 73.2 550 2 S55118 probable membrane
37 30 73.2 769 2 A35848 competence regulat
38 30 73.2 769 2 B69604 two-component sens
39 30 73.2 772 2 T52402 hypothetical prote
40 30 73.2 775 2 S41962 thiamin repressibl
41 30 73.2 947 2 T01238 hypothetical prote
42 30 73.2 1206 2 E86445 hypothetical prote
43 30 73.2 1388 2 A53317 collagen alpha 1(X
44 29 70.7 70 2 S54439 hemSTUV operon pro
45 29 70.7 76 2 T24688 hypothetical prote

ALIGNMENTS

RESULT 1
A56101
collagen alpha 1(XVIII) chain precursor, short splice form - mouse
N;Contains: endostatin
C;Species: Mus musculus (house mouse)
C;Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
C;Accession: A56101; A58371; S72450; S65595
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and ti
tif homologous to rat and Drosophila frizzled proteins.
A;Reference number: A56101; MUID:95181468
A;Accession: A56101
A;Molecule type: mRNA
A;Residues: 1-103 <REH1>
A;Cross-references: GB:U11636; NID:9618427; PIDN:AA052178.1; PID:9618428
R;Rehn, M.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994
A;Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collageno
A;Reference number: A58371; MUID:94240112
A;Accession: A58371
A;Molecule type: mRNA
A;Residues: 1-928 <REH2>
A;Cross-references: GB:L16898; NID:9404754; PIDN:AAA37434.1; PID:9553894
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A;Reference number: S72450
A;Accession: S72450
A;Molecule type: mRNA
A;Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHW>
A;Cross-references: EMBL:L22545; NID:9348968; PIDN:AAA19787.1; PID:9511298
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-
A;Reference number: A58370; MUID:94240111
A;Accession: S65595
A;Molecule type: mRNA
A;Residues: 28-1315 <OHS>
A;Cross-references: EMBL:L22545
C;Comment: Prolines and lysines at the third position of the tripeptide repeating uni
lated and subsequently O-glycosylated.
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in p
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: MGI:Coll8a1
A;Cross-references: MGI:71175
A;Map position: 10:41.0
C;Superfamily: unassigned collagens
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog
F;1-25/Domain: signal sequence #status predicted <SIG>
F;24-235/Region: thrombospondin amino-terminal similarity

F:26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M
 F:327-353/Domain: collagenous #status predicted <CO1>
 F:364-437/Domain: collagenous #status predicted <CO2>
 F:462-583/Domain: collagenous #status predicted <CO3>
 F:607-689/Domain: collagenous #status predicted <CO4>
 F:704-745/Domain: collagenous #status predicted <CO5>
 F:759-831/Domain: collagenous #status predicted <CO6>
 F:842-874/Domain: collagenous #status predicted <CO7>
 F:887-910/Domain: collagenous #status predicted <CO8>
 F:892-894/Region: cell attachment (R-G-D) motif
 F:918-969/Domain: collagenous #status predicted <CO9>
 F:983-1000/Domain: collagenous #status predicted <CO10>
 F:1132-1315/Product: endostatin #status predicted <EST>
 F:1139-1315/Region: multiplexin collagen carboxyl-terminal similarity
 F:126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:172-228/Disulfide bonds: #status predicted
 F:240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 41; DB 2; Length 1315;
 Best Local Similarity 100.0%; Pred. No. 3.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
 |||||

Db 1299 SYIVLCIE 1306

RESULT 2

B56101
 collagen alpha 1(XVIII) chain precursor, long splice form - mouse
 N:Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
 C:Accession: B56101; C56101; S72450; S65595; PN0675; A54072; A58816
 R:Rehn, M.; Pihlajaniemi, T.
 J. Biol. Chem. 270, 4705-4711, 1995
 A:Title: Identification of three N-terminal ends of type XVIII collagen chains and tiss
 tif homologous to rat and Drosophila frizzled proteins.
 A:Reference number: A56101; MUID:95181468
 A:Accession: B56101
 A:Molecule type: mRNA
 A:Residues: 1-562 <REH1>
 A:Cross-references: GB:U11637; NID:g618429; PIDN:AAC52179.1; PID:g618430
 A:Experimental source: splice form clone PE17.24
 A:Accession: C56101
 A:Molecule type: mRNA
 A:Residues: 1-239,487-562 <REH2>
 A:Cross-references: GB:U11637; NID:g618429
 A:Experimental source: splice form clones PE8.1, PE19, PE15.2
 R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
 submitted to the EMBL Data Library, August 1993
 A:Reference number: S72450
 A:Accession: S72450
 A:Molecule type: mRNA
 A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-16
 A:Cross-references: EMBL:L22545; NID:g348968; PIDN:AAA19787.1; PID:g511298
 R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
 Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
 A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
 A:Reference number: A58370; MUID:94240111
 A:Accession: S65595
 A:Molecule type: mRNA
 A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>
 A:Cross-references: EMBL:L22545
 R:Abe, N.; Muragaki, Y.; Yoshioka, H.; Inoue, H.; Ninomiya, Y.
 Biochem. Biophys. Res. Commun. 196, 576-582, 1993
 A:Title: Identification of a novel collagen chain represented by extensive interruptions
 A:Reference number: PN0675; MUID:94059075
 A:Accession: PN0675
 A:Molecule type: mRNA
 A:Residues: 635-1774 <ABE>

R:Rehn, M.; Hintikka, E.; Pihlajaniemi, T.
 J. Biol. Chem. 269, 13929-13935, 1994
 A:Title: Primary structure of the alpha chain of mouse type XVIII collagen, partial
 collagen chain.
 A:Reference number: A54072; MUID:94245707
 A:Accession: A54072
 A:Molecule type: DNA; mRNA
 A:Residues: 1293-1403, 'R', 1405-1774 <REH3>
 A:Cross-references: GB:U03714; NID:g487733; PIDN:AAA20657.1; PID:g487734
 R:O'Reilly, M.S.; Boehm, T.; Shing, Y.; Vasios, G.; Lane, W.S.; Flynn, E.;
 Cell 88, 277-285, 1997
 A:Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
 A:Reference number: A58816; MUID:97160848
 A:Accession: A58816
 A:Molecule type: protein
 A:Residues: 1591-1610 <ORE>
 A:Experimental source: hemangioendothelium cells
 A:Note: Inhibits endothelial cell proliferation
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating uni
 lated and subsequently O-glycosylated.
 C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in p
 C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of
 ay be useful in treating solid tumors.
 C:Genetics:

A:Gene: MGI:Coll8a1
 A:Cross-references: MGI:71175
 A:Map position: 10:41.0
 A:Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 15
 A:Note: the list of introns is incomplete
 C:Superfamily: unassigned collagens
 C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog
 F:1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status p
 F:1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:361-486/Region: frizzled similarity
 F:786-812/Domain: collagenous #status predicted <CO01>
 F:823-896/Domain: collagenous #status predicted <CO02>
 F:921-1042/Domain: collagenous #status predicted <CO03>
 F:1066-1148/Domain: collagenous #status predicted <CO04>
 F:1163-1204/Domain: collagenous #status predicted <CO05>
 F:1218-1290/Domain: collagenous #status predicted <CO06>
 F:1301-1333/Domain: collagenous #status predicted <CO07>
 F:1346-1369/Domain: collagenous #status predicted <CO08>
 F:1351-1353/Region: cell attachment (R-G-D) motif
 F:1377-1428/Domain: collagenous #status predicted <CO09>
 F:1442-1459/Domain: collagenous #status predicted <CO10>
 F:1591-1774/Product: endostatin #status predicted <EST>
 F:1598-1774/Region: multiplexin collagen carboxyl-terminal similarity
 F:354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 41; DB 2; Length 1774;
 Best Local Similarity 100.0%; Pred. No. 5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
 |||||

Db 1758 SYIVLCIE 1765

RESULT 3

A53019
 collagen alpha 1(XVIII) chain - human (fragment)
 N:Contains: endostatin
 C:Species: Homo sapiens (man)
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Mar-2000
 C:Accession: A53019
 R:Oh, S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olse
 Genomics 19, 494-499, 1994
 A:Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and local
 A:Reference number: A53019; MUID:94245237

A;Accession: A53019
A:Molecule type: mRNA
A;Residues: 1-684 <OHA>
A;Cross-references: GB:L22548; NID:g348908; PIDN:AAA51864.1; PID:g562794
A;Note: the cited accession number, L25548, is not in Genbank release 103
A;Note: in the authors' translation, 482-Gly is not shown, residues 483-490 are shifted
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (lated and subsequently O-glycosylated).
C;Comment: Different splice forms of collagen alpha 1(XVII) may be involved in perivascular
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un
ay be useful in treating solid tumors.
C;Genetics:
A;Gene: GDB:COL18A1
A;Cross-references: GDB:138752; OMIM:120328
A;Map position: 21q22.3-21q22.3
C;Superfamily: unassigned collagens
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
F;1-684/Product: collagen alpha 1(XVII) chain (fragment) #status predicted <MAT>
F;1-59/Domain: collagenous (fragment) #status predicted <CO4>
F;74-115/Domain: collagenous #status predicted <CO5>
F;129-201/Domain: collagenous #status predicted <CO6>
F;212-244/Domain: collagenous #status predicted <CO7>
F;257-278/Domain: collagenous #status predicted <CO8>
F;262-264/Region: cell attachment (R-G-D) motif
F;286-340/Domain: collagenous #status predicted <CO9>
F;354-371/Domain: collagenous #status predicted <CO10>
F;502-684/Product: endostatin #status predicted <EST>
F;509-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 92.7%; Score 38; DB 2; Length 684;
Best Local Similarity 87.5%; Pred. No. 8.5;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
:|||||
Db 669 AYIVLCIE 676

RESULT 4
F71498
hypothetical protein CT565 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C;Accession: F71498
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
A;Reference number: A71570; MUID:99000809
A;Accession: F71498
A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-147 <ARN>
A;Cross-references: GB:AE001327; GB:AE001273; NID:g3328999; PIDN:AAC68167.1; PID:g332900
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: CT565

Query Match 78.0%; Score 32; DB 2; Length 147;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
:|||||
Db 123 SFIVLCI 129

RESULT 5
E81656
conserved hypothetical protein TC0854 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C;Accession: E81656
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A;Reference number: A81500; MUID:20150255
A;Accession: E81656
A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-147 <TET>
A;Cross-references: GB:AE002352; GB:AE002160; NID:g7190879; PIDN:AAF39650.1; PID:g719
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0854

Query Match 78.0%; Score 32; DB 2; Length 147;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
:|||||
Db 123 SFIVLCI 129

RESULT 6
D96605
unknown protein F13N6.19 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96605
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: D96605
A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-293 <STO>
A;Cross-references: GB:AE005173; NID:g11024843; PIDN:AAG26928.1; GSPDB:GN00141
C;Genetics:
A;Gene: F13N6.19
A;Map position: 1

Query Match 78.0%; Score 32; DB 2; Length 293;
Best Local Similarity 85.7%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
:|||||
Db 122 SYIALCI 128

RESULT 7
A45031
cysteine-rich fibroblast growth factor receptor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C;Accession: A45031
R;Burrus, L.W.; Zuber, M.E.; Lueddecke, B.A.; Olwin, B.B.
Mol. Cell. Biol. 12, 5600-5609, 1992
A;Title: Identification of a cysteine-rich receptor for fibroblast growth factors.
A;Reference number: A45031; MUID:93078761
A;Accession: A45031
A;Status: preliminary

A:Molecule type: mRNA; protein	
A:Residues: 1-1142 <BUR>	
A:Cross-references: GB:M95766; NID:g211775; PIDN:AAA48769.1; PID:g211776	
A:Experimental source: embryos, brain	
A>Note: sequence extracted from NCBI backbone (NCBIN:119082, NCBIP:119083)	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
C:Keywords: growth factor receptor	
Query Match 78.0%; Score 32; DB 2; Length 1142;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 362 SYLLMCLE 369	
RESULT 8	
S52417	
E-selectin ligand-1 protein - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000	
C:Accession: S52417	
R:Steehmaier, M.; Levinovitz, A.; Isenmann, S.; Borges, E.; Lenter, M.; Kocher, H.P.; Kl	
Nature 373, 615-620, 1995	
A:Title: The E-selectin-ligand ESL-1 is a variant of a receptor for fibroblast growth fa	
A:Reference number: S52417; MUID:95157635	
A:Accession: S52417	
A>Status: preliminary	
A:Molecule type: mRNA	
A:Residues: 1-1175 <STE>	
A:Cross-references: EMBL:X84037; NID:g673435; PIDN:CAA58855.1; PID:g673436	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
Query Match 78.0%; Score 32; DB 2; Length 1175;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 395 SYLLMCLE 402	
RESULT 9	
G85837	
hypothetical protein Z3270 [imported] - Escherichia coli (strain O157:H7, substrain EDL9	
C:Species: Escherichia coli	
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001	
C:Accession: G85837	
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew	
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,	
Nature 409, 529-533, 2001	
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.	
A:Reference number: A85480; MUID:21074935; PMID:11206551	
A:Accession: G85837	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-39 <STO>	
A:Cross-references: GB:AE005174; NID:g12516312; PIDN:AAG57163.1; GSPDB:GN00145; UWGP:Z32	
A:Experimental source: strain O157:H7, substrain EDL933	
C:Genetics:	
A:Gene: Z3270	
Query Match 75.6%; Score 31; DB 2; Length 39;	
Best Local Similarity 57.1%; Pred. No. 17;	
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;	
QY 2 YIVLCIE 8	
: :	
Db 19 FVVLCE 25	
A:Molecule type: mRNA; protein	
A:Residues: 1-1142 <BUR>	
A:Cross-references: GB:M95766; NID:g211775; PIDN:AAA48769.1; PID:g211776	
A:Experimental source: embryos, brain	
A>Note: sequence extracted from NCBI backbone (NCBIN:119082, NCBIP:119083)	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
C:Keywords: growth factor receptor	
Query Match 78.0%; Score 32; DB 2; Length 1142;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 362 SYLLMCLE 369	
RESULT 8	
S52417	
E-selectin ligand-1 protein - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000	
C:Accession: S52417	
R:Steehmaier, M.; Levinovitz, A.; Isenmann, S.; Borges, E.; Lenter, M.; Kocher, H.P.; Kl	
Nature 373, 615-620, 1995	
A:Title: The E-selectin-ligand ESL-1 is a variant of a receptor for fibroblast growth fa	
A:Reference number: S52417; MUID:95157635	
A:Accession: S52417	
A>Status: preliminary	
A:Molecule type: mRNA	
A:Residues: 1-1175 <STE>	
A:Cross-references: EMBL:X84037; NID:g673435; PIDN:CAA58855.1; PID:g673436	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
Query Match 78.0%; Score 32; DB 2; Length 1175;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 395 SYLLMCLE 402	
RESULT 9	
G85837	
hypothetical protein Z3270 [imported] - Escherichia coli (strain O157:H7, substrain EDL9	
C:Species: Escherichia coli	
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001	
C:Accession: G85837	
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew	
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,	
Nature 409, 529-533, 2001	
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.	
A:Reference number: A85480; MUID:21074935; PMID:11206551	
A:Accession: G85837	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-39 <STO>	
A:Cross-references: GB:AE005174; NID:g12516312; PIDN:AAG57163.1; GSPDB:GN00145; UWGP:Z32	
A:Experimental source: strain O157:H7, substrain EDL933	
C:Genetics:	
A:Gene: Z3270	
Query Match 75.6%; Score 31; DB 2; Length 39;	
Best Local Similarity 57.1%; Pred. No. 17;	
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;	
QY 2 YIVLCIE 8	
: :	
Db 19 FVVLCE 25	
A:Molecule type: mRNA; protein	
A:Residues: 1-1142 <BUR>	
A:Cross-references: GB:M95766; NID:g211775; PIDN:AAA48769.1; PID:g211776	
A:Experimental source: embryos, brain	
A>Note: sequence extracted from NCBI backbone (NCBIN:119082, NCBIP:119083)	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
C:Keywords: growth factor receptor	
Query Match 78.0%; Score 32; DB 2; Length 1142;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 362 SYLLMCLE 369	
RESULT 8	
S52417	
E-selectin ligand-1 protein - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000	
C:Accession: S52417	
R:Steehmaier, M.; Levinovitz, A.; Isenmann, S.; Borges, E.; Lenter, M.; Kocher, H.P.; Kl	
Nature 373, 615-620, 1995	
A:Title: The E-selectin-ligand ESL-1 is a variant of a receptor for fibroblast growth fa	
A:Reference number: S52417; MUID:95157635	
A:Accession: S52417	
A>Status: preliminary	
A:Molecule type: mRNA	
A:Residues: 1-1175 <STE>	
A:Cross-references: EMBL:X84037; NID:g673435; PIDN:CAA58855.1; PID:g673436	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
Query Match 78.0%; Score 32; DB 2; Length 1175;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 395 SYLLMCLE 402	
RESULT 9	
G85837	
hypothetical protein Z3270 [imported] - Escherichia coli (strain O157:H7, substrain EDL9	
C:Species: Escherichia coli	
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001	
C:Accession: G85837	
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew	
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,	
Nature 409, 529-533, 2001	
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.	
A:Reference number: A85480; MUID:21074935; PMID:11206551	
A:Accession: G85837	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-39 <STO>	
A:Cross-references: GB:AE005174; NID:g12516312; PIDN:AAG57163.1; GSPDB:GN00145; UWGP:Z32	
A:Experimental source: strain O157:H7, substrain EDL933	
C:Genetics:	
A:Gene: Z3270	
Query Match 75.6%; Score 31; DB 2; Length 39;	
Best Local Similarity 57.1%; Pred. No. 17;	
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;	
QY 2 YIVLCIE 8	
: :	
Db 19 FVVLCE 25	
A:Molecule type: mRNA; protein	
A:Residues: 1-1142 <BUR>	
A:Cross-references: GB:M95766; NID:g211775; PIDN:AAA48769.1; PID:g211776	
A:Experimental source: embryos, brain	
A>Note: sequence extracted from NCBI backbone (NCBIN:119082, NCBIP:119083)	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
C:Keywords: growth factor receptor	
Query Match 78.0%; Score 32; DB 2; Length 1142;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 362 SYLLMCLE 369	
RESULT 8	
S52417	
E-selectin ligand-1 protein - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000	
C:Accession: S52417	
R:Steehmaier, M.; Levinovitz, A.; Isenmann, S.; Borges, E.; Lenter, M.; Kocher, H.P.; Kl	
Nature 373, 615-620, 1995	
A:Title: The E-selectin-ligand ESL-1 is a variant of a receptor for fibroblast growth fa	
A:Reference number: S52417; MUID:95157635	
A:Accession: S52417	
A>Status: preliminary	
A:Molecule type: mRNA	
A:Residues: 1-1175 <STE>	
A:Cross-references: EMBL:X84037; NID:g673435; PIDN:CAA58855.1; PID:g673436	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
Query Match 78.0%; Score 32; DB 2; Length 1175;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 395 SYLLMCLE 402	
RESULT 9	
G85837	
hypothetical protein Z3270 [imported] - Escherichia coli (strain O157:H7, substrain EDL9	
C:Species: Escherichia coli	
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001	
C:Accession: G85837	
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew	
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,	
Nature 409, 529-533, 2001	
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.	
A:Reference number: A85480; MUID:21074935; PMID:11206551	
A:Accession: G85837	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-39 <STO>	
A:Cross-references: GB:AE005174; NID:g12516312; PIDN:AAG57163.1; GSPDB:GN00145; UWGP:Z32	
A:Experimental source: strain O157:H7, substrain EDL933	
C:Genetics:	
A:Gene: Z3270	
Query Match 75.6%; Score 31; DB 2; Length 39;	
Best Local Similarity 57.1%; Pred. No. 17;	
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;	
QY 2 YIVLCIE 8	
: :	
Db 19 FVVLCE 25	
A:Molecule type: mRNA; protein	
A:Residues: 1-1142 <BUR>	
A:Cross-references: GB:M95766; NID:g211775; PIDN:AAA48769.1; PID:g211776	
A:Experimental source: embryos, brain	
A>Note: sequence extracted from NCBI backbone (NCBIN:119082, NCBIP:119083)	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
C:Keywords: growth factor receptor	
Query Match 78.0%; Score 32; DB 2; Length 1142;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 362 SYLLMCLE 369	
RESULT 8	
S52417	
E-selectin ligand-1 protein - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000	
C:Accession: S52417	
R:Steehmaier, M.; Levinovitz, A.; Isenmann, S.; Borges, E.; Lenter, M.; Kocher, H.P.; Kl	
Nature 373, 615-620, 1995	
A:Title: The E-selectin-ligand ESL-1 is a variant of a receptor for fibroblast growth fa	
A:Reference number: S52417; MUID:95157635	
A:Accession: S52417	
A>Status: preliminary	
A:Molecule type: mRNA	
A:Residues: 1-1175 <STE>	
A:Cross-references: EMBL:X84037; NID:g673435; PIDN:CAA58855.1; PID:g673436	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
Query Match 78.0%; Score 32; DB 2; Length 1175;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 395 SYLLMCLE 402	
RESULT 9	
G85837	
hypothetical protein Z3270 [imported] - Escherichia coli (strain O157:H7, substrain EDL9	
C:Species: Escherichia coli	
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001	
C:Accession: G85837	
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew	
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,	
Nature 409, 529-533, 2001	
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.	
A:Reference number: A85480; MUID:21074935; PMID:11206551	
A:Accession: G85837	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-39 <STO>	
A:Cross-references: GB:AE005174; NID:g12516312; PIDN:AAG57163.1; GSPDB:GN00145; UWGP:Z32	
A:Experimental source: strain O157:H7, substrain EDL933	
C:Genetics:	
A:Gene: Z3270	
Query Match 75.6%; Score 31; DB 2; Length 39;	
Best Local Similarity 57.1%; Pred. No. 17;	
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;	
QY 2 YIVLCIE 8	
: :	
Db 19 FVVLCE 25	
A:Molecule type: mRNA; protein	
A:Residues: 1-1142 <BUR>	
A:Cross-references: GB:M95766; NID:g211775; PIDN:AAA48769.1; PID:g211776	
A:Experimental source: embryos, brain	
A>Note: sequence extracted from NCBI backbone (NCBIN:119082, NCBIP:119083)	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
C:Keywords: growth factor receptor	
Query Match 78.0%; Score 32; DB 2; Length 1142;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 362 SYLLMCLE 369	
RESULT 8	
S52417	
E-selectin ligand-1 protein - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000	
C:Accession: S52417	
R:Steehmaier, M.; Levinovitz, A.; Isenmann, S.; Borges, E.; Lenter, M.; Kocher, H.P.; Kl	
Nature 373, 615-620, 1995	
A:Title: The E-selectin-ligand ESL-1 is a variant of a receptor for fibroblast growth fa	
A:Reference number: S52417; MUID:95157635	
A:Accession: S52417	
A>Status: preliminary	
A:Molecule type: mRNA	
A:Residues: 1-1175 <STE>	
A:Cross-references: EMBL:X84037; NID:g673435; PIDN:CAA58855.1; PID:g673436	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
Query Match 78.0%; Score 32; DB 2; Length 1175;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 395 SYLLMCLE 402	
RESULT 9	
G85837	
hypothetical protein Z3270 [imported] - Escherichia coli (strain O157:H7, substrain EDL9	
C:Species: Escherichia coli	
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001	
C:Accession: G85837	
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew	
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,	
Nature 409, 529-533, 2001	
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.	
A:Reference number: A85480; MUID:21074935; PMID:11206551	
A:Accession: G85837	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-39 <STO>	
A:Cross-references: GB:AE005174; NID:g12516312; PIDN:AAG57163.1; GSPDB:GN00145; UWGP:Z32	
A:Experimental source: strain O157:H7, substrain EDL933	
C:Genetics:	
A:Gene: Z3270	
Query Match 75.6%; Score 31; DB 2; Length 39;	
Best Local Similarity 57.1%; Pred. No. 17;	
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;	
QY 2 YIVLCIE 8	
: :	
Db 19 FVVLCE 25	
A:Molecule type: mRNA; protein	
A:Residues: 1-1142 <BUR>	
A:Cross-references: GB:M95766; NID:g211775; PIDN:AAA48769.1; PID:g211776	
A:Experimental source: embryos, brain	
A>Note: sequence extracted from NCBI backbone (NCBIN:119082, NCBIP:119083)	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
C:Keywords: growth factor receptor	
Query Match 78.0%; Score 32; DB 2; Length 1142;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 362 SYLLMCLE 369	
RESULT 8	
S52417	
E-selectin ligand-1 protein - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000	
C:Accession: S52417	
R:Steehmaier, M.; Levinovitz, A.; Isenmann, S.; Borges, E.; Lenter, M.; Kocher, H.P.; Kl	
Nature 373, 615-620, 1995	
A:Title: The E-selectin-ligand ESL-1 is a variant of a receptor for fibroblast growth fa	
A:Reference number: S52417; MUID:95157635	
A:Accession: S52417	
A>Status: preliminary	
A:Molecule type: mRNA	
A:Residues: 1-1175 <STE>	
A:Cross-references: EMBL:X84037; NID:g673435; PIDN:CAA58855.1; PID:g673436	
C:Superfamily: Caenorhabditis elegans hypothetical protein F14E5.2	
Query Match 78.0%; Score 32; DB 2; Length 1175;	
Best Local Similarity 50.0%; Pred. No. 2.1e+02;	
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;	
QY 1 SYIVLCIE 8	
:	
Db 395 SYLLMCLE 402	
RESULT 9	
G85837	
hypothetical protein Z3270 [imported] - Escherichia coli (strain O157:H7, substrain EDL9	
C:Species: Escherichia coli	
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001	
C:Accession: G85837	
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew	
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,	
Nature 409, 529-533, 2001	
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.	
A:Reference number: A85480; MUID:21074935; PMID:11206551	
A:Accession: G85837	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-39 <STO>	
A:Cross-references: GB:AE005174; NID:g12516312; PIDN:AAG57163.1; GSPDB:GN00145; UWGP:Z32	
A:Experimental source: strain O157:H7, substrain EDL933	

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: D84853
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: D84853
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-163 <STO>
A;Cross-references: GB:AE002093; NID:g4567311; PIDN:AAD23722.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g42390
A;Map position: 2

Query Match 75.6%; Score 31; DB 2; Length 163;
Best Local Similarity 62.5%; Pred. No. 60;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
||::||| |
Db 7 SYLLCCE 14

RESULT 13
C97148
hypothetical protein CAC2015 [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: C97148
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C97148
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-174 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK79974.1; PID:g15024998; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2015

Query Match 75.6%; Score 31; DB 2; Length 174;
Best Local Similarity 62.5%; Pred. No. 64;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
|||: |:
Db 80 SYILRCVE 87

RESULT 14
E90114
hypothetical protein orf201 [imported] - Giardia theta nucleomorph
C;Species: nucleomorph Giardia theta
A;Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: E90114
R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei
Nature 410, 1091-1096, 2001
A;Title: The highly reduced genome of an enslaved algal nucleus.
A;Reference number: A99082; MUID:11323671
A;Accession: E90114
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-201 <DOU>
A;Cross-references: GB:AJ010592; NID:g12580769; PIDN:CAC27087.1; GSPDB:GN00151

C;Genetics:
A;Gene: orf201
A;Map position: 2
A;Genome: nucleomorph
C;Keywords: nucleomorph

Query Match 75.6%; Score 31; DB 2; Length 201;
Best Local Similarity 83.3%; Pred. No. 72;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLC 6
|||:|:
Db 132 SYIILC 137

RESULT 15
JC7755
bombinakinin M precursor protein - Chinese red-bellied toad
C;Species: Bombina maxima (Chinese red-bellied toad)
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 14-Dec-2001
C;Accession: JC7755
R;Lai, R.; Liu, H.; Lee, W.H.; Zhang, Y.
Biochem. Biophys. Res. Commun. 286, 259-263, 2001
A;Title: A novel bradykinin-related peptide from skin secretions of toad Bombina max
A;Reference number: JC7755; MUID:21391603; PMID:11500030
A;Contents: Skin secretions
A;Accession: JC7755
A;Molecule type: mRNA
A;Residues: 1-208 <LAI>
A;Cross-references: GB:AY046319
C;Keywords: bradykinin; skin

Query Match 75.6%; Score 31; DB 2; Length 208;
Best Local Similarity 71.4%; Pred. No. 74;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 YIVLCIE 8
:||||:
Db 10 FIVLCLE 16

Search completed: May 31, 2002, 10:28:24
Job time: 166 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:26:28 ; Search time 15.74 Seconds
(without alignments)
15.330 Million cell updates/sec

Title: US-09-589-777C-24
Perfect score: 41
Sequence: 1 SYIVLCIE 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 124036 seqs, 30162252 residues

Total number of hits satisfying chosen parameters: 124036

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	92.7	271	5	US-09-914-277-68
2	33	80.5	134	5	US-09-540-209B-6842
3	33	80.5	568	1	PCT-US02-13142-8459
4	33	80.5	568	6	US-10-128-714-8459
5	31	75.6	502	5	US-09-573-655B-744
6	30	73.2	235	5	US-09-540-209B-6431
7	30	73.2	373	5	US-09-540-209B-7833
8	30	73.2	395	6	US-10-094-080-3
9	30	73.2	417	5	US-09-540-209B-9205
10	30	73.2	728	5	US-09-540-209B-9058
11	30	73.2	1531	5	US-09-970-944-28
12	30	73.2	2174	1	PCT-US02-10824-126
13	29	70.7	51	6	US-10-000-256A-197
14	29	70.7	71	7	US-60-365-384-268
15	29	70.7	86	6	US-10-106-698-4387
16	29	70.7	89	6	US-10-106-698-7796
17	29	70.7	99	5	US-09-620-393B-1341
18	29	70.7	198	5	US-09-540-209B-7063
19	29	70.7	299	1	PCT-US02-09288-4
20	28	68.3	175	6	US-10-106-698-6246
21	28	68.3	296	6	US-10-121-062-32
22	28	68.3	345	6	US-10-126-764-52
23	28	68.3	352	6	US-10-126-764-48
24	28	68.3	353	1	PCT-US02-08413-1
25	28	68.3	353	1	PCT-US02-08413-11
26	28	68.3	353	6	US-10-126-764-2

27	28	68.3	353	6	US-10-126-764-6	Sequence 6, Appli
28	28	68.3	357	6	US-10-126-764-18	Sequence 18, Appl
29	28	68.3	357	6	US-10-126-764-42	Sequence 42, Appl
30	28	68.3	359	6	US-10-126-764-50	Sequence 50, Appl
31	28	68.3	375	6	US-10-126-764-20	Sequence 20, Appl
32	28	68.3	375	6	US-10-126-764-44	Sequence 44, Appl
33	28	68.3	388	6	US-10-126-764-28	Sequence 28, Appl
34	28	68.3	388	6	US-10-126-764-46	Sequence 46, Appl
35	28	68.3	420	5	US-09-591-279-28	Sequence 28, Appl
36	28	68.3	420	5	US-09-591-279-29	Sequence 29, Appl
37	28	68.3	420	5	US-09-591-279-30	Sequence 30, Appl
38	28	68.3	420	5	US-09-591-279-32	Sequence 32, Appl
39	28	68.3	420	5	US-09-591-279-35	Sequence 35, Appl
40	28	68.3	421	5	US-09-591-279-31	Sequence 31, Appl
41	28	68.3	422	6	US-10-126-764-56	Sequence 56, Appl
42	28	68.3	571	5	US-09-540-209B-8124	Sequence 8124, Ap
43	28	68.3	581	6	US-10-041-007-18	Sequence 18, Appl
44	28	68.3	635	5	US-09-540-209B-8038	Sequence 8038, Ap
45	28	68.3	659	5	US-09-573-655B-775	Sequence 775, App

ALIGNMENTS

RESULT 1
US-09-914-277-68
; Sequence 68, Application US/09914277
; GENERAL INFORMATION:
; APPLICANT: Chopp, Michael
; APPLICANT: Wang, Lei
; APPLICANT: Mikkelsen, Tom
; APPLICANT: Dou, Dexian
; TITLE OF INVENTION: AN ANTI-ANGIOGENIC KRINGLE PROTEIN AND ITS MUTANTS
; FILE REFERENCE: 1059.00051
; CURRENT APPLICATION NUMBER: US/09/914,277
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 60/166,176
; PRIOR FILING DATE: 1999-11-18
; PRIOR APPLICATION NUMBER: 60/121,341
; PRIOR FILING DATE: 1999-02-24
; PRIOR APPLICATION NUMBER: 60/121,633
; PRIOR FILING DATE: 1999-02-25
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 68
; LENGTH: 271
; TYPE: PRT
; ORGANISM: homo sapien
US-09-914-277-68

Query Match 92.7%; Score 38; DB 5; Length 271;
Best Local Similarity 87.5%; Pred. No. 4;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 256 AYIVLCIE 263

RESULT 2
US-09-540-209B-6842
; Sequence 6842, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6842
; LENGTH: 134

```
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-6842

Query Match      80.5%; Score 33; DB 5; Length 134;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
   |||||
Db 49 SYCVLCME 56

RESULT 3
PCT-US02-13142-8459
; Sequence 8459, Application PC/TUS0213142
; GENERAL INFORMATION:
; APPLICANT: Elitra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-228
; CURRENT APPLICATION NUMBER: PCT/US02/13142
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8459
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
PCT-US02-13142-8459

Query Match      80.5%; Score 33; DB 1; Length 568;
Best Local Similarity 71.4%; Pred. No. 60;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
   |||||
Db 534 SYIILCL 540

RESULT 4
US-10-128-714-8459
; Sequence 8459, Application US/10128714
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
```

```
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8459
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8459

Query Match      80.5%; Score 33; DB 6; Length 568;
Best Local Similarity 71.4%; Pred. No. 60;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
   |||||
Db 534 SYIILCL 540

RESULT 5
US-09-573-655B-744
; Sequence 744, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptid
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 744
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-744

Query Match      75.6%; Score 31; DB 5; Length 502;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIVLCI 7
   |||||
Db 191 SYVVFCI 197

RESULT 6
US-09-540-209B-6431
; Sequence 6431, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6431
; LENGTH: 235
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-6431

Query Match      73.2%; Score 30; DB 5; Length 235;
Best Local Similarity 71.4%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YIVLCIE 8
```



```
RESULT 11
US-09-970-944-28
; Sequence 28, Application US/09970944
; GENERAL INFORMATION:
; APPLICANT: Herriman, John L
; APPLICANT: Rastelli, Luca
; APPLICANT: Shimkets, Richard A
; TITLE OF INVENTION: Novel Proteins and Nucleic Acids Encoding Same and
; TITLE OF INVENTION: Antibodies Directed Against these Proteins
; FILE REFERENCE: 21402-138
; CURRENT APPLICATION NUMBER: US/09/970,944
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: 60/237,862
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-970-944-28
```

```
Query Match 73.2%; Score 30; DB 5; Length 1531;
Best Local Similarity 62.5%; Pred. No. 4.8e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
|| |||
Db 1104 SYACLCVE 1111
```

```
RESULT 12
PCT-US02-10824-126
; Sequence 126, Application PC/TUS0210824
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Prostate Cancer Expression Profiles
; FILE REFERENCE: 9U 206 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/10824
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: US 60/281,732
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/281,731
; PRIOR FILING DATE: 2001-04-06
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 126
; LENGTH: 2174
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-10824-126
```

```
Query Match 73.2%; Score 30; DB 1; Length 2174;
Best Local Similarity 62.5%; Pred. No. 6.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
:|:| | |
Db 1370 AYVVLCP E 1377
```

```
RESULT 13
US-10-000-256A-197
; Sequence 197, Application US/10000256A
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and
```

```
; FILE REFERENCE: DEX-0259
; CURRENT APPLICATION NUMBER: US/10/000,256A
; CURRENT FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: 60/244,782
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 197
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-000-256A-197
```

```
Query Match 70.7%; Score 29; DB 6; Length 51;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 YIVLCI 7
|:| | |
Db 38 YLILCI 43
```

```
RESULT 14
US-60-365-384-268
; Sequence 268, Application US/60365384
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ghosh, Malabika
; APPLICANT: Zhao, Qing A.
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Weng, Gezhi
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: 814
; CURRENT APPLICATION NUMBER: US/60/365,384
; CURRENT FILING DATE: 2002-03-14
; NUMBER OF SEQ ID NOS: 666
; SOFTWARE: pt_FL-genes Version 6.0
; SEQ ID NO 268
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-365-384-268
```

```
Query Match 70.7%; Score 29; DB 7; Length 71;
Best Local Similarity 71.4%; Pred. No. 57;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YIVLCIE 8
|| |||
Db 38 YIQLCLE 44
```

```
RESULT 15
US-10-106-698-4387
; Sequence 4387, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
```

```
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4387
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (22)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (54)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (64)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (70)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-10-106-698-4387
```

```
Query Match          70.7%; Score 29; DB 6; Length 86;
Best Local Similarity 50.0%; Pred. No. 67;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 SYIVLCIE 8
   |::||:|
Db 24 SFTLCLE 31
```

Search completed: May 31, 2002, 10:32:17
Job time: 349 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:25:53 ; Search time 205.32 Seconds
(without alignments)
13.714 Million cell updates/sec

Title: US-09-589-777C-24
Perfect score: 41
Sequence: 1 SYIVLCIE 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues
Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main.*
1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	8	19	US-09-589-777C-24
2	41	100.0	48	1	PCT-US00-12063-21
3	41	100.0	48	17	US-09-353-333-21
4	41	100.0	50	23	US-09-958-489-4
5	41	100.0	85	19	US-09-589-774-5
6	41	100.0	184	1	PCT-US00-12063-4
7	41	100.0	184	1	PCT-US98-24950-10

8	41	100.0	184	13	US-08-975-424-10	Sequence 10, Appl
9	41	100.0	184	17	US-09-353-333-4	Sequence 4, Appl
10	41	100.0	184	17	US-09-383-315-18	Sequence 18, Appl
11	41	100.0	184	19	US-09-589-777A-2	Sequence 2, Appl
12	41	100.0	184	19	US-09-589-777C-2	Sequence 2, Appl
13	41	100.0	185	24	US-10-036-869-36	Sequence 36, Appl
14	41	100.0	191	19	US-09-561-005-13	Sequence 13, Appl
15	41	100.0	191	19	US-09-561-499-13	Sequence 13, Appl
16	41	100.0	191	19	US-09-561-526-13	Sequence 13, Appl
17	41	100.0	191	19	US-09-562-245-13	Sequence 13, Appl
18	41	100.0	191	23	US-09-998-831-13	Sequence 13, Appl
19	41	100.0	207	19	US-09-586-625-71	Sequence 71, Appl
20	41	100.0	207	24	US-10-080-797-3	Sequence 3, Appl
21	41	100.0	218	1	PCT-US98-24950-14	Sequence 14, Appl
22	41	100.0	218	13	US-08-975-424-14	Sequence 14, Appl
23	41	100.0	580	1	PCT-US98-24950-16	Sequence 16, Appl
24	41	100.0	580	13	US-08-975-424-16	Sequence 16, Appl
25	38	92.7	27	1	PCT-US00-12063-79	Sequence 79, Appl
26	38	92.7	27	17	US-09-353-333-79	Sequence 79, Appl
27	38	92.7	84	19	US-09-589-774-6	Sequence 6, Appl
28	38	92.7	139	23	US-09-978-531-12	Sequence 12, Appl
29	38	92.7	170	15	US-09-171-607-1	Sequence 1, Appl
30	38	92.7	178	17	US-09-325-116-2	Sequence 2, Appl
31	38	92.7	182	17	US-09-325-116-1	Sequence 1, Appl
32	38	92.7	182	19	US-09-561-005-14	Sequence 14, Appl
33	38	92.7	182	19	US-09-561-499-14	Sequence 14, Appl
34	38	92.7	182	19	US-09-561-526-14	Sequence 14, Appl
35	38	92.7	182	19	US-09-562-245-14	Sequence 14, Appl
36	38	92.7	182	23	US-09-998-831-14	Sequence 14, Appl
37	38	92.7	183	1	PCT-US00-12063-2	Sequence 2, Appl
38	38	92.7	183	1	PCT-US00-12063-23	Sequence 23, Appl
39	38	92.7	183	1	PCT-US98-24950-18	Sequence 18, Appl
40	38	92.7	183	13	US-08-975-424-18	Sequence 18, Appl
41	38	92.7	183	15	US-09-164-057-76	Sequence 76, Appl
42	38	92.7	183	17	US-09-353-333-2	Sequence 2, Appl
43	38	92.7	183	17	US-09-353-333-23	Sequence 23, Appl
44	38	92.7	183	17	US-09-383-315-4	Sequence 4, Appl
45	38	92.7	183	18	US-09-457-018-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-589-777C-24
; Sequence 24, Application US/09589777C
; GENERAL INFORMATION:
; APPLICANT: Sukhatme, Vikas P.
; TITLE OF INVENTION: Anti-Angiogenic Peptides and Methods of
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: 1440.1023-011
; CURRENT APPLICATION NUMBER: US/09/589,777C
; CURRENT FILING DATE: 2000-06-08
; PRIOR APPLICATION NUMBER: PCT/US98/26057
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 60/108,536
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: US 60/082,663
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: US 60/067,888
; PRIOR FILING DATE: 1997-12-08
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-589-777C-24

Query Match 100.0%; Score 41; DB 19; Length 8;
Best Local Similarity 100.0%; Pred. No. 3.2e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 1 SYIVLCIE 8

RESULT 2
PCT-US00-12063-21
; Sequence 21, Application PC/TUS0012063
; GENERAL INFORMATION:
; APPLICANT: The Burnham Institute
; TITLE OF INVENTION: Antiangiogenic Endostatin Peptides, Endostatin Variants
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: FP-LJ 4148
; CURRENT APPLICATION NUMBER: PCT/US00/12063
; CURRENT FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: US 09/353,333
; PRIOR FILING DATE: 1999-07-14
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Mus musculus
PCT-US00-12063-21

Query Match 100.0%; Score 41; DB 1; Length 48;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 37 SYIVLCIE 44

RESULT 3
US-09-353-333-21
; Sequence 21, Application US/09353333A
; GENERAL INFORMATION:
; APPLICANT: Vuori, Kristiina
; TITLE OF INVENTION: Antiangiogenic Endostatin Peptides, Endostatin Variants
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: P-LJ 3557
; CURRENT APPLICATION NUMBER: US/09/353,333A
; CURRENT FILING DATE: 1999-07-14
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-353-333-21

Query Match 100.0%; Score 41; DB 17; Length 48;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 37 SYIVLCIE 44

RESULT 4
US-09-958-489-4
; Sequence 4, Application US/09958489
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITA' DEGLI STUDI DI MILANO
; APPLICANT: UNIVERSITA' DEGLI STUDI DI FIRENZE
; TITLE OF INVENTION: POLYPEPTIDES WITH ANTIANGIOGENIC ACTIVITY
; FILE REFERENCE: UNIVERSITA'
; CURRENT APPLICATION NUMBER: US/09/958,489

; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (39)
; OTHER INFORMATION: T-BUTYL
; OTHER INFORMATION: Description of Artificial Sequence: polypeptides
; OTHER INFORMATION: homologous to endostatin
; NAME/KEY: DISULFID
; LOCATION: (1)..(31)
; OTHER INFORMATION: intramolecular disulfide bond
US-09-958-489-4

Query Match 100.0%; Score 41; DB 23; Length 50;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 34 SYIVLCIE 41

RESULT 5
US-09-589-774-5
; Sequence 5, Application US/09589774
; GENERAL INFORMATION:
; APPLICANT: Sukhatme, Vikas P.
; TITLE OF INVENTION: Restin and Methods of Use Thereof
; FILE REFERENCE: 1440.1014011
; CURRENT APPLICATION NUMBER: US/09/589,774
; CURRENT FILING DATE: 2000-06-08
; PRIOR APPLICATION NUMBER: PCT/US98/26058
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: 60/108,536
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 60/082,663
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/067,888
; PRIOR FILING DATE: 1997-12-08
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-589-774-5

Query Match 100.0%; Score 41; DB 19; Length 85;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 69 SYIVLCIE 76

RESULT 6
PCT-US00-12063-4
; Sequence 4, Application PC/TUS0012063
; GENERAL INFORMATION:
; APPLICANT: The Burnham Institute
; TITLE OF INVENTION: Antiangiogenic Endostatin Peptides, Endostatin Variants
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: FP-LJ 4148
; CURRENT APPLICATION NUMBER: PCT/US00/12063
; CURRENT FILING DATE: 2000-05-02

; PRIOR APPLICATION NUMBER: US 09/353,333
; PRIOR FILING DATE: 1999-07-14
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
PCT-US00-12063-4

Query Match 100.0%; Score 41; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 168 SYIVLCIE 175

RESULT 7
PCT-US98-24950-10
; Sequence 10, Application PC/TUS9824950
; GENERAL INFORMATION:
; APPLICANT: Genetix Pharmaceuticals, Inc.
; TITLE OF INVENTION: ANTI-ANGIOGENIC GENE THERAPY VECTORS AND
; TITLE OF INVENTION: THEIR USE IN TREATING ANGIOGENESIS-RELATED DISEASES
; FILE REFERENCE: 50033/002W01
; CURRENT APPLICATION NUMBER: PCT/US98/24950
; CURRENT FILING DATE: 1998-12-29
; EARLIER APPLICATION NUMBER: 08/975,424
; EARLIER FILING DATE: 1997-11-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
PCT-US98-24950-10

Query Match 100.0%; Score 41; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 168 SYIVLCIE 175

RESULT 8
US-08-975-424-10
; Sequence 10, Application US/08975424
; GENERAL INFORMATION:
; APPLICANT: LeBoulch, Philipp
; TITLE OF INVENTION: ANTI-ANGIOGENIC GENE THERAPY VECTORS AND
; TITLE OF INVENTION: THEIR USE IN TREATING ANGIOGENESIS-RELATED DISEASES
; FILE REFERENCE: 50033/002001
; CURRENT APPLICATION NUMBER: US/08/975,424
; CURRENT FILING DATE: 1997-11-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-975-424-10

Query Match 100.0%; Score 41; DB 13; Length 184;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 168 SYIVLCIE 175

RESULT 9
US-09-353-333-4
; Sequence 4, Application US/09353333A
; GENERAL INFORMATION:
; APPLICANT: Vuori, Kristiina
; TITLE OF INVENTION: Antiangiogenic Endostatin Peptides, Endostatin Variants
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: P-LJ 3557
; CURRENT APPLICATION NUMBER: US/09/353,333A
; CURRENT FILING DATE: 1999-07-14
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-353-333-4

Query Match 100.0%; Score 41; DB 17; Length 184;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 168 SYIVLCIE 175

RESULT 10
US-09-383-315-18
; Sequence 18, Application US/09383315A
; GENERAL INFORMATION:
; APPLICANT: Lo, Kin-Ming
; APPLICANT: Li, Yue
; APPLICANT: Gillies, Stephen D
; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as
; TITLE OF INVENTION: Immunofusins
; FILE REFERENCE: LEX-006
; CURRENT APPLICATION NUMBER: US/09/383,315A
; CURRENT FILING DATE: 1999-08-25
; EARLIER APPLICATION NUMBER: US 60/097,883
; EARLIER FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-383-315-18

Query Match 100.0%; Score 41; DB 17; Length 184;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 168 SYIVLCIE 175

RESULT 11
US-09-589-777A-2
; Sequence 2, Application US/09589777A
; GENERAL INFORMATION:
; APPLICANT: Sukhatme, Vikas P.
; TITLE OF INVENTION: Anti-Angiogenic Peptides and Methods of
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: 1440.1023-011

; CURRENT APPLICATION NUMBER: US/09/589,777A
; CURRENT FILING DATE: 2000-06-08
; PRIOR APPLICATION NUMBER: PCT/US98/26057
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: US 60/108,536
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: US 60/082,663
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: US 60/067,888
; PRIOR FILING DATE: 1997-12-07
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-589-777A-2

Query Match 100.0%; Score 41; DB 19; Length 184;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 168 SYIVLCIE 175

RESULT 12
US-09-589-777C-2
; Sequence 2, Application US/09589777C
; GENERAL INFORMATION:
; APPLICANT: Sukhatme, Vikas P.
; TITLE OF INVENTION: Anti-Angiogenic Peptides and Methods of
; FILE REFERENCE: 1440.1023-011
; CURRENT APPLICATION NUMBER: US/09/589,777C
; CURRENT FILING DATE: 2000-06-08
; PRIOR APPLICATION NUMBER: PCT/US98/26057
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 60/108,536
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: US 60/082,663
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: US 60/067,888
; PRIOR FILING DATE: 1997-12-08
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-589-777C-2

Query Match 100.0%; Score 41; DB 19; Length 184;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 168 SYIVLCIE 175

RESULT 13
US-10-036-869-36
; Sequence 36, Application US/10036869
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; THERAPY
; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/036,869
; FILING DATE: 29-Nov-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 185 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-10-036-869-36

Query Match 100.0%; Score 41; DB 24; Length 185;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 169 SYIVLCIE 176

RESULT 14
US-09-561-005-13
; Sequence 13, Application US/09561005
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002585
; CURRENT APPLICATION NUMBER: US/09/561,005
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-005-13

Query Match 100.0%; Score 41; DB 19; Length 191;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 169 SYIVLCIE 176

Db 175 SYIVLCIE 182

RESULT 15
US-09-561-499-13
; Sequence 13, Application US/09561499
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/09/561,499
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-499-13

Query Match 100.0%; Score 41; DB 19; Length 191;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
|||||||
Db 175 SYIVLCIE 182

Search completed: May 31, 2002, 10:31:55
Job time: 362 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:25:18 ; Search time 21.9 Seconds
(without alignments)
8.923 Million cell updates/sec

Title: US-09-589-777C-24

Perfect score: 41

Sequence: 1 SYIVLCIE 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	41	100.0	185	3	US-08-985-526-36	Sequence 36, Appl
2	41	100.0	191	4	US-09-561-500-13	Sequence 13, Appl
3	41	100.0	191	4	US-09-561-108-13	Sequence 13, Appl
4	41	100.0	195	1	US-08-159-784-2	Sequence 2, Appl
5	38	92.7	178	4	US-09-315-689-5	Sequence 5, Appl
6	38	92.7	182	4	US-09-561-500-14	Sequence 14, Appl
7	38	92.7	182	4	US-09-561-108-14	Sequence 14, Appl
8	38	92.7	182	4	US-09-315-689-3	Sequence 3, Appl
9	38	92.7	183	4	US-09-206-059-2	Sequence 2, Appl
10	30	73.2	154	3	US-09-191-647-10	Sequence 10, Appl
11	30	73.2	154	4	US-09-540-245A-10	Sequence 10, Appl
12	30	73.2	154	4	US-09-540-153-10	Sequence 10, Appl
13	30	73.2	191	1	US-08-159-784-3	Sequence 3, Appl
14	30	73.2	254	2	US-08-948-616-9	Sequence 9, Appl
15	30	73.2	254	2	US-09-193-510-9	Sequence 9, Appl
16	30	73.2	254	4	US-09-368-402-9	Sequence 9, Appl
17	30	73.2	307	2	US-08-948-616-3	Sequence 3, Appl
18	30	73.2	307	2	US-09-193-510-3	Sequence 3, Appl
19	30	73.2	307	4	US-09-368-402-3	Sequence 3, Appl
20	30	73.2	395	1	US-08-723-938-3	Sequence 3, Appl
21	30	73.2	395	2	US-09-080-538-3	Sequence 3, Appl
22	30	73.2	420	4	US-09-008-271A-4	Sequence 4, Appl
23	30	73.2	420	4	US-08-974-691-8	Sequence 8, Appl
24	30	73.2	470	2	US-08-724-394A-10	Sequence 10, Appl
25	30	73.2	775	2	US-08-714-070A-1	Sequence 1, Appl
26	29	70.7	113	2	US-08-466-860-8	Sequence 8, Appl
27	29	70.7	113	3	US-08-472-040A-8	Sequence 8, Appl

28	29	70.7	113	4	US-08-276-776-8	Sequence 8, Appl
29	29	70.7	113	4	US-08-471-209-8	Sequence 8, Appl
30	29	70.7	343	2	US-08-878-989-5	Sequence 5, Appl
31	29	70.7	343	4	US-09-272-796-5	Sequence 5, Appl
32	29	70.7	417	4	US-09-347-833-8	Sequence 8, Appl
33	29	70.7	1503	4	US-08-976-255-14	Sequence 14, Appl
34	28	68.3	318	4	US-09-387-574-2	Sequence 2, Appl
35	28	68.3	318	4	US-09-668-096-2	Sequence 2, Appl
36	28	68.3	353	3	US-08-984-288-2	Sequence 2, Appl
37	28	68.3	353	4	US-09-224-426-4	Sequence 4, Appl
38	28	68.3	353	4	US-09-478-601-4	Sequence 4, Appl
39	28	68.3	353	4	US-09-478-602-4	Sequence 4, Appl
40	28	68.3	400	5	PCT-US95-16472-2	Sequence 2, Appl
41	28	68.3	402	3	US-08-602-809-2	Sequence 2, Appl
42	28	68.3	422	4	US-09-224-426-2	Sequence 2, Appl
43	28	68.3	422	4	US-09-478-601-2	Sequence 2, Appl
44	28	68.3	422	4	US-09-478-602-2	Sequence 2, Appl
45	28	68.3	468	4	US-09-355-115-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-985-526-36
; Sequence 36, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/608,845
; APPLICATION NUMBER: 16-JUL-1996
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 185 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-985-526-36

Query Match 100.0%; Score 41; DB 3; Length 185;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYIVLCIE 8

|||||||

Db 169 SYIVLCIE 176

```

RESULT 2
US-09-561-500-13
; Sequence 13, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-500-13

Query Match      100.0%; Score 41; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SYIVLCIE 8
      |||||
Db      175 SYIVLCIE 182

RESULT 3
US-09-561-108-13
; Sequence 13, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-108-13

Query Match      100.0%; Score 41; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SYIVLCIE 8
      |||||
Db      175 SYIVLCIE 182

RESULT 4
US-08-159-784-2
; Sequence 2, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
```

```

; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
US-08-159-784-2

Query Match      100.0%; Score 41; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SYIVLCIE 8
      |||||
Db      179 SYIVLCIE 186

RESULT 5
US-09-315-689-5
; Sequence 5, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-5

Query Match      92.7%; Score 38; DB 4; Length 178;
Best Local Similarity 87.5%; Pred. No. 2.4;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 SYIVLCIE 8
      :|||
```

Db 164 AYIVLCIE 171

RESULT 6

US-09-561-500-14

; Sequence 14, Application US/09561500

; Patent No. 6342219

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Rolf A. Brekken

; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF

; FILE REFERENCE: 4001.002500

; CURRENT APPLICATION NUMBER: US/09/561,500

; CURRENT FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/131,432

; PRIOR FILING DATE: 1999-04-28

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 14

; LENGTH: 182

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

; OTHER INFORMATION: PEPTIDE

US-09-561-500-14

Query Match 92.7%; Score 38; DB 4; Length 182;

Best Local Similarity 87.5%; Pred. No. 2.5;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8

:|||||

Db 168 AYIVLCIE 175

RESULT 7

US-09-561-108-14

; Sequence 14, Application US/09561108

; Patent No. 6342221

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Rolf A. Brekken

; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF

; FILE REFERENCE: 4001.002584

; CURRENT APPLICATION NUMBER: US/09/561,108

; CURRENT FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/131,432

; PRIOR FILING DATE: 1999-04-28

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 14

; LENGTH: 182

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

; OTHER INFORMATION: PEPTIDE

US-09-561-108-14

Query Match 92.7%; Score 38; DB 4; Length 182;

Best Local Similarity 87.5%; Pred. No. 2.5;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8

:|||||

Db 168 AYIVLCIE 175

RESULT 8

US-09-315-689-3

; Sequence 3, Application US/09315689

; Patent No. 6346510

; GENERAL INFORMATION:

; APPLICANT: Folkman, Judah

; APPLICANT: O'Reilly, Michael

; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions

; FILE REFERENCE: 05213-0229

; CURRENT APPLICATION NUMBER: US/09/315,689

; CURRENT FILING DATE: 1999-05-20

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 182

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-315-689-3

Query Match 92.7%; Score 38; DB 4; Length 182;

Best Local Similarity 87.5%; Pred. No. 2.5;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8

:|||||

Db 168 AYIVLCIE 175

RESULT 9

US-09-206-059-2

; Sequence 2, Application US/09206059

; Patent No. 6201104

; GENERAL INFORMATION:

; APPLICANT: MacDonald, Nicholas

; APPLICANT: Sim, Kim Lee

; TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and

; TITLE OF INVENTION: Proteins and Methods of Use

; FILE REFERENCE: 05213-0370

; CURRENT APPLICATION NUMBER: US/09/206,059

; CURRENT FILING DATE: 1998-12-04

; NUMBER OF SEQ ID NOS: 80

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 183

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-206-059-2

Query Match 92.7%; Score 38; DB 4; Length 183;

Best Local Similarity 87.5%; Pred. No. 2.5;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8

:|||||

Db 168 AYIVLCIE 175

RESULT 10

US-09-191-647-10

; Sequence 10, Application US/09191647

; Patent No. 6046015

; GENERAL INFORMATION:

; APPLICANT: Goodman, Corey

; APPLICANT: Kid, Thomas

; APPLICANT: Brose, Katja

; APPLICANT: Tessier-Lavigne, Marc

; TITLE OF INVENTION: Modulating Robo: Ligand Interactions

; FILE REFERENCE: B98-031-3

; CURRENT APPLICATION NUMBER: US/09/191,647

; CURRENT FILING DATE: 1998-11-13

; EARLIER APPLICATION NUMBER: 60/065,544

; EARLIER FILING DATE: 1997-11-14

; EARLIER APPLICATION NUMBER: 60/081,057

; EARLIER FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 154
; TYPE: PRT
; ORGANISM: mouse
US-09-191-647-10

Query Match 73.2%; Score 30; DB 3; Length 154;
Best Local Similarity 62.5%; Pred. No. 70;
Matches 5; Conservative 1; Mismatches 2; Indels 2; Gaps 0;

QY 1 SYIVLCIE 8
|| ||:|
Db 44 SYACLCVE 51

RESULT 11
US-09-540-245A-10
; Sequence 10, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 154
; TYPE: PRT
; ORGANISM: mouse
US-09-540-245A-10

Query Match 73.2%; Score 30; DB 4; Length 154;
Best Local Similarity 62.5%; Pred. No. 70;
Matches 5; Conservative 1; Mismatches 2; Indels 2; Gaps 0;

QY 1 SYIVLCIE 8
|| ||:|
Db 44 SYACLCVE 51

RESULT 12
US-09-540-153-10
; Sequence 10, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,153
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/191,647
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10
; LENGTH: 154
; TYPE: PRT
; ORGANISM: mouse
US-09-540-153-10

Query Match 73.2%; Score 30; DB 4; Length 154;
Best Local Similarity 62.5%; Pred. No. 70;
Matches 5; Conservative 1; Mismatches 2; Indels 2; Gaps 0;

QY 1 SYIVLCIE 8
|| ||:|
Db 44 SYACLCVE 51

RESULT 13
US-08-159-784-3
; Sequence 3, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
US-08-159-784-3

Query Match 73.2%; Score 30; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVLCIE 8
|||||
Db 177 IVLCIE 182

RESULT 14
US-08-948-616-9
; Sequence 9, Application US/08948616
; Patent No. 5840539

GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,616
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1223894
US-08-948-616-9

Query Match 73.2%; Score 30; DB 2; Length 254;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
||:||||
Db 69 SYMMLCTE 76

RESULT 15
US-09-193-510-9
Sequence 9, Application US/09193510
Patent No. 5981226
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/193,510
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/948,616
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1223894
US-09-193-510-9
Query Match 73.2%; Score 30; DB 2; Length 254;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 SYIVLCIE 8
||:||||
Db 69 SYMMLCTE 76

Search completed: May 31, 2002, 10:27:50
Job time: 152 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:23:23 ; Search time 51.8 Seconds
(without alignments)
17.154 Million cell updates/sec

Title: US-09-589-777C-24

Perfect score: 41

Sequence: 1 SYIVLCIE 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802:*

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	8	AA18407	C-terminus of anti
2	41	100.0	48	AA19808	Murine endostatin
3	41	100.0	50	AA19858	Antiangiogenic pen
4	41	100.0	184	AA18409	Endostatin protein
5	41	100.0	184	AA19868	Murine endostatin
6	41	100.0	184	AA19868	Murine endostatin
7	41	100.0	184	AA19868	Murine endostatin
8	41	100.0	185	AA19868	Anti-angiogenic en
9	41	100.0	191	AA19868	Murine endostatin
10	41	100.0	207	AA19868	Murine endostatin
11	41	100.0	207	AA19868	Murine endostatin

12	41	100.0	218	20	AA19869	Murine gene therap
13	41	100.0	580	20	AA19869	Murine gene therap
14	41	100.0	684	20	AA19869	Mouse alpha1 (XVII
15	41	100.0	1288	18	AA19869	Mouse alpha-1 coll
16	41	100.0	1288	20	AA19869	Mouse alpha-1 (XVI
17	38	92.7	27	22	AA19869	Human endostatin p
18	38	92.7	178	21	AA19869	Alternate human en
19	38	92.7	178	22	AA19869	Human Endostatin(T
20	38	92.7	179	22	AA19869	Human Endostatin(T
21	38	92.7	180	22	AA19869	Human Endostatin(T
22	38	92.7	181	22	AA19869	Human Endostatin(T
23	38	92.7	182	21	AA19869	Human endostatin.
24	38	92.7	182	21	AA19869	Human endostatin p
25	38	92.7	182	21	AA19869	Human endostatin p
26	38	92.7	182	22	AA19869	Human endostatin(T
27	38	92.7	183	20	AA19869	Human endostatin p
28	38	92.7	183	20	AA19869	SEQ ID 76 of WO991
29	38	92.7	183	21	AA19869	Amino acid sequenc
30	38	92.7	183	21	AA19869	Human endostatin p
31	38	92.7	183	21	AA19869	Human angiogenesis
32	38	92.7	183	21	AA19869	Human angiogenesis
33	38	92.7	183	22	AA19869	Human Endostatin(T
34	38	92.7	183	22	AA19869	Human endostatin S
35	38	92.7	183	22	AA19869	Human endostatin p
36	38	92.7	184	22	AA19869	Human vascular end
37	38	92.7	193	21	AA19869	Human HMW endostat
38	38	92.7	195	21	AA19869	Human HMW endostat
39	38	92.7	216	21	AA19869	Amino acid sequenc
40	38	92.7	271	21	AA19869	A human angiogenes
41	38	92.7	684	18	AA19869	Human alpha-1 coll
42	38	92.7	684	20	AA19869	Human alpha1 (XVII
43	38	92.7	737	22	AA19869	Novel human diagno
44	38	92.7	1301	20	AA19869	Human alpha-1 (XVI
45	38	92.7	1336	20	AA19869	Human collagen 18

ALIGNMENTS

RESULT 1

AA18407

ID AAY18407 standard; peptide; 8 AA.

XX AAY18407;

XX

DT 24-AUG-1999 (first entry)

DE C-terminus of anti-angiogenic peptide EM1.

XX

KW EM1; anti-angiogenic peptide; endostatin; angiogenesis-dependent cancer;

KW benign tumour; rheumatoid arthritis; psoriasis; ocular angiogenesis;

KW Osler-Webber Syndrome; myocardial angiogenesis; angiofibroma; cancer;

KW plaque neovascularisation; telangiectasia; atherosclerosis; scleroderma;

KW dialysis graft vascular access stenosis; renal cancer; therapy.

OS Mus sp.

XX

PN WO9929855-A1.

PD

PD 17-JUN-1999.

XX

PF 08-DEC-1998; 98WO-US26057.

XX

PR 16-NOV-1998; 98US-0108536.

PR 08-DEC-1997; 97US-0067888.

PR 22-APR-1998; 98US-0082663.

XX

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX

PI Sukhatme VP;

XX

DR WPI; 1999-385604/32.

XX

PT Mutant endostatin having anti-angiogenic activity
XX
PS Claim 1; Page 71; 105pp; English.
XX
CC This sequence represents the C-terminus from the mutant endostatin (EM)
CC of the invention, which has anti-angiogenic activity, and is designated
CC EM1. Compositions comprising EM1 or fusion proteins comprising EM1, are
CC useful for treating diseases characterised by angiogenic activity, such
CC as angiogenesis-dependent cancers, benign tumours, rheumatoid arthritis,
CC psoriasis, ocular angiogenesis, Osler-Webber syndrome, myocardial
CC angiogenesis, plaque neovascularisation, telangiectasia, haemophiliac
CC joints, angiofibroma, wound granulation, intestinal adhesions,
CC atherosclerosis, scleroderma, hypertrophic scars, cat scratch disease,
CC Helicobacter pylori ulcers, dialysis graft vascular access stenosis,
CC contraception and obesity. In particular, the diseases treatable by EM1
CC comprise cancer, especially renal cancer. The methods provide a means for
CC introducing EM1 into mammalian cells via gene therapy, for production of
CC EM1 via recombinant means, as well as recombinant production of the EM1
CC protein. EM1 performs as well or better than whole endostatin. Use
CC of EM1 is advantageous for treatment of angiogenic diseases in that
CC increasingly smaller peptides are more potent on a weight basis, and may
CC be able to better penetrate tissues.
XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 41; DB 20; Length 8;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 1 syivlcie 8

RESULT 2
AAB49808
ID AAB49808 standard; Protein; 48 AA.

XX
AC AAB49808;
XX
DT 02-MAR-2001 (first entry)
XX
DE Murine endostatin peptide fragment SEQ ID NO: 21.
XX
KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;
KW cancer; inflammation; angiogenesis-dependent disease.

OS Mus musculus.
XX
PN WO200067771-A1.
XX
PD 16-NOV-2000.
XX
PF 02-MAY-2000; 2000WO-US12063.
XX
PR 06-MAY-1999; 99US-0132907.
PR 14-JUL-1999; 99US-0353333.
XX
PA (BURN-) BURNHAM INST.
XX
PI Vuori K;
XX
WPI; 2001-040937/05.

XX Endostatin peptide comprising at least four endostatin amino acid
PT residues are e.g. angiogenesis inhibitors for treating cancer and
PT diabetic retinopathy -
XX
PS Claim 10; Page 124-125; 146pp; English.
XX
CC The present invention provides endostatin peptides which can be used in
CC the modulation of angiogenesis. This is useful in the treatment of

CC cancers, inflammation, rheumatoid arthritis, chronic articular
CC rheumatism, psoriasis, disorders associated with inopportune invasion of
CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
CC of prematurity, macular degeneration, corneal graft rejection,
CC retrolental fibroplasia, rubeosis, capillary proliferation in
CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
CC diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophiliac joints and wound
CC granulation. In addition, the peptides can be used as birth control
CC agents.
XX
SQ Sequence 48 AA;

Query Match 100.0%; Score 41; DB 22; Length 48;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 37 syivlcie 44

RESULT 3
AAB35587
ID AAB35587 standard; peptide; 50 AA.

XX
AC AAB35587;
XX
DT 14-FEB-2001 (first entry)
XX
DE Antiangiogenic pentapeptide IV.
XX
KW Antiangiogenic; angiogenesis; cancer; endostatin.
XX

OS Synthetic.

XX WO200063249-A1.

XX 26-OCT-2000.

XX 11-APR-2000; 2000WO-EP03236.

XX 15-APR-1999; 99IT-MI00777.

XX (UYMI-) UNIV MILANO.

XX (UYFI-) UNIV FIRENZE.

XX Chillemi F, Francescato P, Ziche M;

XX WPI; 2001-007005/01.

XX Polypeptides derived from endostatin exhibiting antiangiogenic activity
PT useful for treatment of angiogenesis-dependent tumours -

XX Claim 5; Page 18; 28pp; English.

XX The present invention describes a number of peptides derived from
CC endostatin which exhibit antiangiogenic activity. These may be used in
CC the treatment of cancer. The present sequence is one of the peptides of
CC the invention.
XX

SQ Sequence 50 AA;

Query Match 100.0%; Score 41; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8

Db 34 syivlcie 41

```

RESULT 4
AAY18409
ID AAY18409 standard; Protein; 184 AA.
XX
AC AAY18409;
XX
DT 24-AUG-1999 (first entry)
XX
DE Endostatin protein sequence.
XX
KW EM1; anti-angiogenic peptide; endostatin; angiogenesis-dependent cancer;
KW benign tumour; rheumatoid arthritis; psoriasis; ocular angiogenesis;
KW Osler-Webber Syndrome; myocardial angiogenesis; angiofibroma; cancer;
KW plaque neovascularisation; telangiectasia; atherosclerosis; scleroderma;
KW dialysis graft vascular access stenosis; renal cancer; therapy.
XX
OS Mus sp.
XX
PN WO9929855-A1.
XX
PD 17-JUN-1999.
XX
PF 08-DEC-1998; 98WO-US26057.
XX
PR 16-NOV-1998; 98US-0108536.
PR 08-DEC-1997; 97US-0067888.
PR 22-APR-1998; 98US-0082663.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Sukhatme VP;
XX
DR WPI; 1999-385604/32.
DR N-PSDB; AAX79949.
XX
PT Mutant endostatin having anti-angiogenic activity
XX
PS Claim 31; Fig 2; 105pp; English.
XX
CC This sequence is the mouse endostatin. The invention relates to a
CC the mutant endostatin (EM), which has anti-angiogenic activity, and is
CC designated EM1. Compositions comprising EM1 or fusion proteins comprising
CC EM1, are useful for treating diseases characterised by angiogenic
CC activity, such as angiogenesis-dependent cancers, benign tumours,
CC rheumatoid arthritis, psoriasis, ocular angiogenesis, Osler-Webber
CC Syndrome, myocardial angiogenesis, plaque neovascularisation,
CC telangiectasia, haemophilic joints, angiofibroma, wound granulation,
CC intestinal adhesions, atherosclerosis, scleroderma, hypertrophic scars,
CC cat scratch disease, Helicobacter pylori ulcers, dialysis graft vascular
CC access stenosis, contraception and obesity. In particular, the diseases
CC treatable by EM1 comprise cancer, especially renal cancer. The methods
CC provide a means for introducing EM1 into mammalian cells via gene
CC therapy, for production of EM1 via recombinant means, as well as
CC recombinant production of the EM1 protein. EM1 performs as well or better
CC than whole endostatin. Use of EM1 is advantageous for treatment of
CC angiogenic diseases in that increasingly smaller peptides are more potent
CC on a weight basis, and may be able to better penetrate tissues.
XX
SQ Sequence 184 AA;

Query Match 100.0%; Score 41; DB 20; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 168 syivlcie 175

RESULT 5
AAY08689
```

```

ID XX AAY08689 standard; Protein; 184 AA.
XX AC AAY08689;
XX DT 10-AUG-1999 (first entry)
XX DE Murine endostatin protein fragment.
XX KW Plasminogen; murine; angiotatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina.
XX OS Mus sp.
XX PN WO9926480-A1.
XX PD 03-JUN-1999.
XX PF 20-NOV-1998; 98WO-US24950.
XX PR 20-NOV-1997; 97US-0975424.
XX (GENE-) GENETIX PHARM INC.
XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX Bachelot T, Leboulch P, Pawliuk RJ;
XX WPI; 1999-357696/30.
XX DR N-PSDB; AAX77715.
XX PT Anti-angiogenic gene therapy vectors
XX PS Disclosure; Fig 6; 83pp; English.
XX CC This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiotatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.
XX Sequence 184 AA;

Query Match 100.0%; Score 41; DB 20; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 168 syivlcie 175

RESULT 6
AAY70258
ID AAY70258 standard; Protein; 184 AA.
XX AC AAY70258;
XX DT 06-JUN-2000 (first entry)
XX DE Murine angiogenesis inhibitor, endostatin.
XX KW Murine; immunoglobulin Fc fragment; endostatin; immunofusin;
KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
KW vasotropic; vulnery; treatment; antiarteriosclerosis; tumour;
```

KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW wound granulation; keloid scar; gene therapy.
XX
OS Mus musculus.
XX WO200011033-A2.
XX PD 02-MAR-2000.
XX PF 25-AUG-1999; 99WO-US19329.
XX PR 25-AUG-1998; 98US-0097883.
XX PA (LEXI-) LEXINGEN PHARM CORP.
XX PI Lo K, Li Y, Gillies SD;
XX WPI; 2000-237616/20.
XX N-PSDB; AAZ51299.
XX
PT Novel fusion protein of angiotatin or endostatin and an immunoglobulin
PT FC region, useful for treating conditions mediated by angiogenesis,
PT such as rheumatoid arthritis, tumors and macular degeneration -
XX
XX Example 5; Pages 48-49; 68pp; English.
XX
CC The patent discloses a DNA molecule encoding a fusion protein comprising
CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis
CC inhibitor selected from angiotatin, endostatin, a plasminogen fragment
CC having angiotatin activity, a collagen XVIII fragment having endostatin
CC activity, or combinations of them. The fusion protein (immunofusin) is
CC used to inhibit angiogenesis and to treat diseases or conditions mediated
CC by angiogenesis. Conditions that may be treated include solid tumours,
CC blood born tumours, tumour metastasis, benign tumours including
CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
CC e.g. diabetic retinopathy, retinopathy of prematurity, macular
CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental
CC fibroplasia, rubeosis and Osler-Webber syndrome; myocardial angiogenesis,
CC plaque neovascularisation, telangiectasia, haemophiliac joints,
CC angiofibroma, wound granulation, and excessive or abnormal stimulation of
CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used
CC in gene therapy. The present sequence is a murine
CC endostatin used in the construction of immunofusin containing murine
CC immunoglobulin Fc fragment.
XX
SQ Sequence 184 AA;

Query Match 100.0%; Score 41; DB 21; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 168 syivlcie 175

RESULT 7
AAB49380
ID AAB49380 standard; Protein; 184 AA.
XX AAB49380;

XX 02-MAR-2001 (first entry)
XX Murine endostatin SEQ ID NO: 4.
DE Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;
XX cancer; inflammation; angiogenesis-dependent disease.
KW

XX Mus musculus.
OS WO200067771-A1.
XX PN 16-NOV-2000.
XX PD 02-MAY-2000; 2000WO-US12063.
XX PF 06-MAY-1999; 99US-0132907.
XX PR 14-JUL-1999; 99US-0353333.
XX PA (BURN-) BURNHAM INST.
XX XX
XX Vuori K;
XX WPI; 2001-040937/05.
XX N-PSDB; AAC88290.
XX
PT Endostatin peptide comprising at least four endostatin amino acid
PT residues are e.g. angiogenesis inhibitors for treating cancer and
PT diabetic retinopathy -
XX
XX Disclosure; Fig 1; 146pp; English.
XX
CC The present invention provides endostatin peptides which can be used in
CC the modulation of angiogenesis. This is useful in the treatment of
CC cancers, inflammation, rheumatoid arthritis, chronic articular
CC rheumatism, psoriasis, disorders associated with inopportune invasion of
CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
CC of prematurity, macular degeneration, corneal graft rejection,
CC retrolental fibroplasia, rubeosis, capillary proliferation in
CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
CC diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophiliac joints and wound
CC granulation. In addition, the peptides can be used as birth control
CC agents.
XX
SQ Sequence 184 AA;

Query Match 100.0%; Score 41; DB 22; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | |
Db 168 syivlcie 175

RESULT 8
AAY06197
ID AAY06197 standard; Protein; 185 AA.
XX AAY06197;
AC AAY06197;

XX 16-AUG-1999 (first entry)
XX Anti-angiogenic endostatin peptide.
DE

XX Anti-angiogenic; carrier:DNA complex; tumour; gene therapy; human;
KW endostatin; melanoma; lung cancer; colon cancer; brain cancer;
KW breast cancer.

XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 36 /note= "encoded by CAG"
FT Misc-difference 37 /note= "encoded by CAA"
FT Misc-difference 39 /note= "deduced sequence from nucleotide sequence

FT has an Ala residue between residues 39 and
FT 40 of this sequence"

FT Misc-difference 76 /note= "encoded by AAG"
FT Misc-difference 118 /note= "encoded by AAG"
FT Misc-difference 162 /note= "encoded by AAA"
FT Misc-difference 168 /note= "encoded by AAC"
FT Misc-difference 185 /note= "encoded by AAA"
FT
XX
PN EP921193-A1.
XX
PD 09-JUN-1999.
XX
PF 07-JAN-1998; 98EP-0100135.
XX
PR 05-DEC-1997; 97US-0985526.
XX
PA (MIXS/) MIXSON A J.
XX
PI Mixson AJ;
XX
DR WPI; 1999-315406/27.
DR N-PSDB; AAX58740.
XX
PT Inhibition of growth of solid tumors
XX
PS Disclosure; Page 38; 46pp; English.
XX
CC The present sequence represents an anti-angiogenic endostatin
CC peptide. The invention provides a carrier:DNA complex that comprises
CC DNA (see AAX58725-42) encoding an anti-angiogenic protein or peptide,
CC such as the present sequence, the complex being deliverable to
CC the site of a tumour in vivo, and additionally comprises regulatory
CC elements for expressing the anti-angiogenic DNA in a tumour or
CC tumour vasculature. The complex may also include DNA encoding a
CC tumour suppressor protein, especially p53. The carrier is a
CC liposome, cationic polymer, micelle, microsphere, virus, viral
CC component, or a combination of these, and administration is by
CC intravenous or intratumoral injection. The complexes are useful in
CC gene therapy for inhibition of tumour growth. The types of tumors
CC which may be treated include solid tumors such as melanomas and
CC tumors in the lung, colon, brain and breast.
XX
SQ Sequence 185 AA;

Query Match 100.0%; Score 41; DB 20; Length 185;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 169 syivlcie 176
IIIIIIII

RESULT 9
AAB28398
ID AAB28398 standard; Protein; 191 AA.
XX
AC AAB28398;
XX
DT 19-FEB-2001 (first entry)
XX
DE Murine endostatin.
XX
KW Murine; endostatin; cytostatic; antiproliferative;
KW vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;
KW cancer; vascularised solid tumour.
XX

OS Mus sp.
XX
PN WO200064946-A2.
XX
PD 02-NOV-2000.
XX
PF 28-APR-2000; 2000WO-US11367.
XX
PR 28-APR-1999; 99US-0131432.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Thorpe PE, Brekken RA;
XX
DR WPI; 2000-687317/67.
DR N-PSDB; AAC67777.
XX
PT Immunogenic composition for the treatment and diagnosis of cancer
PT comprises an anti-VEGF (vascular endothelial growth factor) antibody
PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -
XX
PS Example 10; Page 290-291; 298pp; English.
XX
CC The present invention relates to anti-Vascular Endothelial Growth Factor
CC (VEGF) antibodies that bind to the same epitope as the monoclonal
CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to
CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF
CC receptor VEGFR1. The present sequence is murine endostatin. Endostatin
CC may be conjugated onto the anti-VEGF antibodies of the present invention.
CC The anti-VEGF antibodies of the present invention are useful for the
CC treatment and diagnosis of cancer, especially vascularised solid tumours.
XX
SQ Sequence 191 AA;

Query Match 100.0%; Score 41; DB 21; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
Db 175 syivlcie 182
IIIIIIII

RESULT 10
AAE02031
ID AAE02031 standard; Protein; 207 AA.
XX
AC AAE02031;
XX
DT 31-JUL-2001 (first entry)
XX
DE Murine endostatin fused to N-terminal secretion signal.
XX
KW Murine; endostatin; fusion protein; nucleotide-binding domain; NBD;
KW ligand-binding domain; LBD; transcription regulating domain; TRD; cancer;
KW zinc finger protein; ZFP; ligand-activated transcriptional regulator;
KW gene regulation; gene therapy; cell proliferative disorder; psoriasis;
KW pemphigus vulgaris; Behcet's syndrome; lipid histiocytosis.
XX
OS Mus sp.
XX
PN WO200130843-A1.
XX
PD 03-MAY-2001.
XX
PF 23-OCT-2000; 2000WO-EP10430.
XX
PR 25-OCT-1999; 99US-0433042.
PR 02-JUN-2000; 2000US-0586625.
XX
PA (NOVS) NOVARTIS AG.
PA (SCRI) SCRIPPS RES INST.

XX Barbas CF, Kadan M, Beerli R;
XX WPI; 2001-308618/32.
DR N-PSDB; AAD06108.
XX
PT New fusion protein containing nucleotide-binding and ligand-binding
PT domains, useful e.g. in gene therapy of cancer, provides
PT ligand-activated control of gene expression .
XX
PS Example 19; Page 209; 218pp; English.
XX
CC The invention relates to fusion protein comprising a nucleotide-binding
CC domain (NBD), a ligand-binding domain (LBD) of an intracellular receptor
CC (ICR) and a transcription regulating domain (TRD). NBD is a polydactyl
CC zinc finger protein (ZFP), or a modular part of it, that interacts
CC specifically with a contiguous sequence of at least 3 nucleotides. The
CC fusion protein functions as a ligand-activated transcriptional regulator.
CC The fusion protein and the nucleic acid encoding it, are used to regulate
CC gene expression, particularly in gene therapy for treating malignant
CC cell proliferative diseases (e.g. colon cancer, prostate cancer,
CC renal-cell carcinoma) and non-malignant cell proliferative
CC diseases (e.g. psoriasis, pemphigus vulgaris, Behcet's syndrome and
CC lipid histiocytosis). The fusion protein and its DNA are also useful for
CC treating diseases caused by viruses in humans/plants, genetic and/or
CC acquired diseases. The fusion protein can be designed to target any
CC selected gene (endogenous or exogenous), and can be made to have
CC different selectivity or specificity for endogenous or exogenous ligands.
CC The present sequence is murine endostatin fused to an N-terminal
CC secretion signal. The corresponding cDNA sequence was used in the
CC construction of Left end shuttle plasmids containing regulatable
CC transgene cassettes for evaluation of Cys2-His2 Zinc finger DNA binding
CC domain (DBD)-Oestrogen receptor (ER) LBD regulators.
XX
SQ Sequence 207 AA;

Query Match 100.0%; Score 41; DB 22; Length 207;
Best Local Similarity 100.0%; Pred. NO. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
|||
Db 191 syivlcie 198

RESULT 11
AAB71930
ID AAB71930 standard; Protein; 207 AA.

XX AAB71930;
AC
XX 10-MAY-2001 (first entry)
XX Murine endostatin attached to Ig-kappa leader sequence.
DE
XX Mouse; endostatin; antitumour; cytostatic; antiarthritic; antipsoriatic;
KW antidiabetic; ophthalmological; gene therapy; angiogenic inhibitor;
KW adenoviral vector; diabetic retinopathy; cardiovascular disease;
KW arthritis; psoriasis; cerebral oedema; intravascular coagulopathy;
KW lymphoma; leukaemia; immunoglobulin; Ig; Ig-kappa.
XX
OS Mus sp.
XX WO200112830-A1.
PN
XX 22-FEB-2001.
PD
XX 11-AUG-2000; 2000WO-EP07865.
PF
XX 13-AUG-1999; 99US-0373938.
PR
XX (NOVS) NOVARTIS AG.

PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX Hallenbeck PL, Chen CT;
PI
XX WPI; 2001-202871/20.
DR N-PSDB; AAF60336.
XX
PT Adenoviral vector for treating tumors and disorders associated with
PT angiogenesis, such as cancer, arthritis, and psoriasis, comprises a DNA
PT sequence encoding an angiogenic inhibitor, particularly endostatin .
XX
PS Example 1; Fig 1B; 59pp; English.
XX
CC The nucleotide sequence encoding this protein was used in the
CC construction of an adenoviral vector which includes a DNA sequence
CC encoding endostatin. The adenoviral vector is useful for expressing
CC endostatin in a mammalian cell such as an A549 or Hep3B cell. It is
CC useful for treating other diseases and disorders associated with
CC angiogenesis, such as neovascular diseases of the eye, including diabetic
CC retinopathy, cardiovascular disease, arthritis, psoriasis, cerebral
CC oedema and intravascular coagulopathy (Kasabach-Merritt syndrome). The
CC vector inhibits, prevents or destroys the growth of tumours by
CC preventing the formation of blood vessels in tumours, such as lymphoma
CC and leukaemia.
XX
SQ Sequence 207 AA;

Query Match 100.0%; Score 41; DB 22; Length 207;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
|||
Db 191 syivlcie 198

RESULT 12
AAY08691
ID AAY08691 standard; Protein; 218 AA.

XX AAY08691;
AC
XX 10-AUG-1999 (first entry)
XX Murine gene therapy peptide construct SP-Flag-Endo.
DE
XX Plasminogen; murine; angiotatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.
XX
OS Mus sp.
OS Synthetic.
XX WO9926480-A1.
PN
XX 03-JUN-1999.
PD
XX 20-NOV-1998; 98WO-US24950.
PF
XX 20-NOV-1997; 97US-0975424.
PR
XX (GENE-) GENETIX PHARM INC.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Bachelot T, Leboulch P, Pawliuk RJ;
XX WPI; 1999-357696/30.
DR N-PSDB; AAX77717.
XX
PT Anti-angiogenic gene therapy vectors
XX Example 1; Page 69; 83pp; English.
PS

XX This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiostatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.
XX
SQ Sequence 218 AA;

Query Match 100.0%; Score 41; DB 20; Length 218;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | | | | |
Db 202 syivlcie 209

RESULT 13
AAY08692
ID AAY08692 standard; Protein; 580 AA.

XX AC AAY08692;
XX DT 10-AUG-1999 (first entry)
XX DE Murine gene therapy peptide construct SP-K1-K2-K3-K4-Flag-Endo.

XX KW Plasminogen; murine; angiostatin; endostatin; gene therapy; vector;
XX anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
XX tumour growth; solid tumour; diabetic retinopathy; retina; construct.

XX OS Mus sp.
XX Synthetic.
XX PN W09926480-A1.
XX PD 03-JUN-1999.

XX PF 20-NOV-1998; 98WO-US24950.
XX PR 20-NOV-1997; 97US-0975424.
XX PA (GENE-) GENETIX PHARM INC.
XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX PI Bachelot T, Leboulch P, Pawliuk RJ;
XX WPI; 1999-357696/30.
XX DR N-PSDB; AAX77718.

XX PT Anti-angiogenic gene therapy vectors
XX Example 1; Page 72-74; 83pp; English.

XX This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiostatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide

CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.
XX Sequence 580 AA;

Query Match 100.0%; Score 41; DB 20; Length 580;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8
| | | | | | | |
Db 564 syivlcie 571

RESULT 14
AAY25114
ID AAY25114 standard; Protein; 684 AA.

XX AC AAY25114;
XX DT 25-AUG-1999 (first entry)
XX DE Mouse alpha1 (XVIII) collagen protein.

XX KW Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;
XX anti-angiogenic; heparin binding domain; receptor binding domain; mimic;
XX alpha-helix A domain; carbohydrate recognition domain; CRD domain;
XX treatment; angiogenesis; tumour; murine.

XX OS Mus sp.
XX PN W09931616-A1.
XX PD 24-JUN-1999-

XX PF 16-DEC-1998; 98WO-US26783.
XX PR 16-DEC-1997; 97US-0069727.

XX PA (HARD) HARVARD COLLEGE.
XX PI Hohenester E, Olsen BR, Sasaki T, Timpl R;
XX WPI; 1999-395243/33.
XX PT Identifying mimetics of mammalian endostatin

XX PS Disclosure; Fig 5A-C; 75pp; English.

XX This invention describes a novel method for identifying mimetics of
CC mammalian endostatin. The method comprises identifying a compound
CC having atomic coordinates with non-trivial similarity to selected
CC coordinates of atoms of a mammalian endostatin involves (a) providing
CC a library of atomic coordinates of compounds in a library of candidate
CC compounds, (b) comparing the library of atomic coordinates to the
CC selected coordinates of a mammalian endostatin and (c) selecting from the
CC library at least one candidate compound on the basis of selection
CC criteria which include similarities between the atomic coordinates of the
CC selected candidate compound and the atomic coordinates of the mammalian
CC endostatin. The invention also describes the use of an anti-angiogenic
CC fragment of endostatin comprising a domain selected from a heparin
CC binding domain, a receptor binding domain, and exposed on alpha-helix A
CC domain, and a carbohydrate recognition domain (CRD) domain. The methods
CC can be used for designing and selecting endostatin mimics. The compounds
CC identified can be used for treating undesired angiogenesis, e.g. tumours.
CC This sequence represents mouse alpha1(XVIII) collagen which is used in
CC the description of the method.

XX Sequence 684 AA;

Query Match 100.0%; Score 41; DB 20; Length 684;

Best Local Similarity 100.0%; Pred. No. 6;			
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 SYIVLCIE 8		
Db	669 syivlcie 676		
RESULT 15			
AAW26328			
ID	AAW26328 standard; Protein; 1288 AA.		
XX	AAW26328;		
AC			
XX			
DT	19-NOV-1997 (first entry)		
XX			
DE	Mouse alpha-1 collagen (XVIII).		
XX			
KW	Alpha-1 collagen; type XVIII collagen; cartilage degeneration.		
XX			
OS	Mus musculus.		
Key Location/Qualifiers			
FH	Peptide	303..308	
FT		/label= GYGX'Y'_motif	
FT	Peptide	309..314	
FT		/label= GYGX'Y'_motif	
FT	Peptide	315..320	
FT		/label= GYGX'Y'_motif	
FT	Peptide	321..326	
FT		/label= GYGX'Y'_motif	
FT	Peptide	337..342	
FT		/label= GYGX'Y'_motif	
FT	Peptide	343..348	
FT		/label= GYGX'Y'_motif	
FT	Peptide	349..354	
FT		/label= GYGX'Y'_motif	
FT	Peptide	355..360	
FT		/label= GYGX'Y'_motif	
FT	Peptide	361..366	
FT		/label= GYGX'Y'_motif	
FT	Peptide	367..372	
FT		/label= GYGX'Y'_motif	
FT	Peptide	373..378	
FT		/label= GYGX'Y'_motif	
FT	Peptide	379..384	
FT		/label= GYGX'Y'_motif	
FT	Peptide	385..390	
FT		/label= GYGX'Y'_motif	
FT	Peptide	396..401	
FT		/label= GYGX'Y'_motif	
FT	Peptide	402..407	
FT		/label= GYGX'Y'_motif	
FT	Peptide	435..440	
FT		/label= GYGX'Y'_motif	
FT	Peptide	441..446	
FT		/label= GYGX'Y'_motif	
FT	Peptide	447..452	
FT		/label= GYGX'Y'_motif	
FT	Peptide	453..458	
FT		/label= GYGX'Y'_motif	
FT	Peptide	459..464	
FT		/label= GYGX'Y'_motif	
FT	Peptide	470..475	
FT		/label= GYGX'Y'_motif	
FT	Peptide	476..481	
FT		/label= GYGX'Y'_motif	
FT	Peptide	482..487	
FT		/label= GYGX'Y'_motif	
FT	Peptide	488..493	
FT		/label= GYGX'Y'_motif	
FT	Peptide	494..499	
FT		/label= GYGX'Y'_motif	

FT	Peptide	500..505	
FT		/label= GYGX'Y'_motif	
FT	Peptide	506..511	
FT		/label= GYGX'Y'_motif	
FT	Peptide	512..517	
FT		/label= GYGX'Y'_motif	
FT	Peptide	518..523	
FT		/label= GYGX'Y'_motif	
FT	Peptide	524..529	
FT		/label= GYGX'Y'_motif	
FT	Peptide	530..535	
FT		/label= GYGX'Y'_motif	
FT	Peptide	536..541	
FT		/label= GYGX'Y'_motif	
FT	Peptide	542..547	
FT		/label= GYGX'Y'_motif	
FT	Peptide	548..553	
FT		/label= GYGX'Y'_motif	
FT	Peptide	580..585	
FT		/label= GYGX'Y'_motif	
FT	Peptide	586..591	
FT		/label= GYGX'Y'_motif	
FT	Peptide	592..597	
FT		/label= GYGX'Y'_motif	
FT	Peptide	598..603	
FT		/label= GYGX'Y'_motif	
FT	Peptide	604..609	
FT		/label= GYGX'Y'_motif	
FT	Peptide	610..615	
FT		/label= GYGX'Y'_motif	
FT	Peptide	616..621	
FT		/label= GYGX'Y'_motif	
FT	Peptide	622..627	
FT		/label= GYGX'Y'_motif	
FT	Peptide	628..633	
FT		/label= GYGX'Y'_motif	
FT	Peptide	634..639	
FT		/label= GYGX'Y'_motif	
FT	Peptide	640..665	
FT		/label= GYGX'Y'_motif	
FT	Peptide	657..662	
FT		/label= GYGX'Y'_motif	
FT	Peptide	677..682	
FT		/label= GYGX'Y'_motif	
FT	Peptide	683..688	
FT		/label= GYGX'Y'_motif	
FT	Peptide	689..694	
FT		/label= GYGX'Y'_motif	
FT	Peptide	695..700	
FT		/label= GYGX'Y'_motif	
FT	Peptide	707..712	
FT		/label= GYGX'Y'_motif	
FT	Peptide	713..718	
FT		/label= GYGX'Y'_motif	
FT	Peptide	735..740	
FT		/label= GYGX'Y'_motif	
FT	Peptide	741..746	
FT		/label= GYGX'Y'_motif	
FT	Peptide	747..752	
FT		/label= GYGX'Y'_motif	
FT	Peptide	759..764	
FT		/label= GYGX'Y'_motif	
FT	Peptide	765..770	
FT		/label= GYGX'Y'_motif	
FT	Peptide	771..776	
FT		/label= GYGX'Y'_motif	
FT	Peptide	787..792	
FT		/label= GYGX'Y'_motif	
FT	Peptide	793..798	
FT		/label= GYGX'Y'_motif	
FT	Peptide	799..804	
FT		/label= GYGX'Y'_motif	
FT	Peptide	815..820	

FT /label= GYGX'Y'_motif
FT 821..826
FT /label= GYGX'Y'_motif
FT 827..832
FT /label= GYGX'Y'_motif
FT 833..838
FT /label= GYGX'Y'_motif
FT 839..844
FT /label= GYGX'Y'_motif
FT 845..850
FT /label= GYGX'Y'_motif
FT 863..868
FT /label= GYGX'Y'_motif
FT 869..874
FT /label= GYGX'Y'_motif
FT 875..880
FT /label= GYGX'Y'_motif
FT 891..896
FT /label= GYGX'Y'_motif
FT 897..902
FT /label= GYGX'Y'_motif
FT 903..908
FT /label= GYGX'Y'_motif
FT 911..916
FT /label= GYGX'Y'_motif
FT 917..922
FT /label= GYGX'Y'_motif
FT 928..933
FT /label= GYGX'Y'_motif
FT 934..939
FT /label= GYGX'Y'_motif
FT 956..961
FT /label= GYGX'Y'_motif
FT 962..967
FT /label= GYGX'Y'_motif
FT 968..973
FT /label= GYGX'Y'_motif
FT 1126..1131
FT /label= GYGX'Y'_motif
FT 1145..1150
FT /label= GYGX'Y'_motif
FT 1193..1198
FT /label= GYGX'Y'_motif
XX

US5643783-A.

01-JUL-1997.

'01-DEC-1993; 93US-0159784.

01-DEC-1993; 93US-0159784.

(HARD) HARVARD COLLEGE.

Oh SP, Olsen BR;

WPI; 1997-350247/32.

N-PSDB; AAT84485.

XX Nucleic acid encoding human alpha-1 collagen - for production of
PT recombinant alpha-1 collagen, for use in the treatment of cartilage
PT degeneration

XX PS Disclosure; Fig 2; 35pp; English.

Query Match 100.0%; Score 41; DB 18; Length 1288;
Best Local Similarity 100.0%; Pred. NO. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIVLCIE 8

Db 1272 syivlcie 1279

Search completed: May 31, 2002, 10:27:22
Job time: 239 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:33:06 ; Search time 41.34 Seconds
(without alignments)
37.562 Million cell updates/sec

Title: US-09-589-777C-25
Perfect score: 45
Sequence: 1 NSFMTFSK 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_19:

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	160	11 Q9CRT2	Q9crt2 mus musculu
2	45	100.0	184	11 Q9JK63	Q9jk63 mus musculu
3	45	100.0	226	11 Q9QZD2	Q9qzd2 rattus norv
4	45	100.0	1140	11 Q61434	Q61434 mus musculu
5	45	100.0	1774	11 Q62001	Q62001 mus musculu
6	37	82.2	349	5 Q965U4	Q965u4 caenorhabdi
7	34	75.6	128	5 Q9XX30	Q9xx30 caenorhabdi
8	34	75.6	263	5 Q9N312	Q9n312 caenorhabdi
9	34	75.6	296	12 Q56767	Q56767 human cytom
10	34	75.6	604	8 Q950Y9	Q950y9 tetrahymena
11	33	73.3	108	16 Q97T55	Q97t55 streptococc
12	33	73.3	300	5 Q18743	Q18743 caenorhabdi
13	33	73.3	792	10 Q9FGD4	Q9fgd4 arabidopsis
14	33	73.3	825	10 Q9AUQ2	Q9auq2 oryza sativ
15	32	71.1	183	5 Q9NAE6	Q9nae6 caenorhabdi
16	32	71.1	203	16 Q97KI0	Q97ki0 clostridium

17	32	71.1	461	10 Q92RG9	Q9zrg9 oryza sativ
18	32	71.1	582	10 Q03865	Q03865 zea mays (m
19	32	71.1	593	10 Q9MAB4	Q9mab4 arabidopsis
20	32	71.1	655	10 Q9C620	Q9c620 arabidopsis
21	32	71.1	690	3 Q9P8I3	Q9p8i3 candida tro
22	32	71.1	1056	5 Q96935	Q96935 plasmodium
23	32	71.1	3933	5 Q97239	Q97239 plasmodium
24	32	71.1	4083	3 Q9C1M7	Q9clm7 ashbya goss
25	32	71.1	7829	5 Q18559	Q18559 caenorhabdi
26	31	68.9	265	16 Q9CEM6	Q9cem6 lactococcus
27	31	68.9	271	10 Q38711	Q38711 arachis hyp
28	31	68.9	312	2 Q50603	Q50603 staphylococ
29	31	68.9	312	2 Q53746	Q53746 staphylococ
30	31	68.9	312	9 Q80066	Q80066 staphylococ
31	31	68.9	368	16 Q92XN2	Q92xn2 rhizobium m
32	31	68.9	405	5 Q95XV0	Q95xv0 caenorhabdi
33	31	68.9	542	4 Q9BRG0	Q9brg0 homo sapien
34	31	68.9	585	16 Q97FM7	Q97fm7 clostridium
35	31	68.9	598	17 Q27162	Q27162 methanother
36	31	68.9	614	10 Q9FKB6	Q9fkb6 arabidopsis
37	31	68.9	720	16 Q97FP9	Q97fp9 clostridium
38	31	68.9	740	11 Q9D332	Q9d332 mus musculu
39	31	68.9	836	5 Q62201	Q62201 caenorhabdi
40	31	68.9	1003	5 Q9W0L5	Q9w0l5 drosophila
41	31	68.9	1042	5 Q9GRV9	Q9grv9 caenorhabdi
42	30	66.7	108	16 Q9PME4	Q9pme4 campylobact
43	30	66.7	179	12 Q9Q8H8	Q9q8h8 myxoma viru
44	30	66.7	191	12 Q91MP0	Q91mp0 lumpy skin
45	30	66.7	248	10 Q43375	Q43375 arachis hyp

ALIGNMENTS

RESULT 1
Q9CRT2
ID Q9CRT2
AC Q9CRT2;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PROCOLLAGEN, TYPE XVIII, ALPHA 1 (FRAGMENT).
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momota N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK014292; BAB29249.1; -.
HSSP; P39061; 1KOE.

DR MGD; MGI:88451; Coll18a1.
FT NON_TER 1
SQ SEQUENCE 160 AA; 17725 MW; 60F853D777C375D2 CRC64;

Query Match 100.0%; Score 45; DB 11; Length 160;
Best Local Similarity 100.0%; Pred. No. 0.068;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||
Db 152 NSFMTSFSK 160

RESULT 2
Q9JK63 ID Q9JK63 PRELIMINARY; PRT; 184 AA.
AC Q9JK63;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ENDOSTATIN (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHINESE KUNMING;
RA Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;
RT "Anticancer treatment of targeted fusion protein delivery to tumor
neovasculature";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF257775; AAF69009.1; -.
DR HSSP; P39061; 1KOE.
FT NON_TER 1
FT NON_TER 184
SQ SEQUENCE 184 AA; 20376 MW; AC06F9D8D103412A CRC64;

Query Match 100.0%; Score 45; DB 11; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||
Db 176 NSFMTSFSK 184

RESULT 3
Q9QZD2 ID Q9QZD2 PRELIMINARY; PRT; 226 AA.
AC Q9QZD2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE COLLAGEN XVIII (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=20227226; PubMed=10766159;
RA Perletti G., Concari P., Giardini R., Marras E., Piccinini F.,
RA Folkman J., Chen L.;
RT "Antitumor activity of endostatin against carcinogen-induced rat
primary mammary tumors";
RL Cancer Res. 60:1793-1796(2000).
DR EMBL; AF189709; AAF00975.1; -.
DR HSSP; P39061; 1KOE.
FT NON_TER 1

Query Match 100.0%; Score 45; DB 11; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||
Db 176 NSFMTSFSK 184

RESULT 3
Q9QZD2 ID Q9QZD2 PRELIMINARY; PRT; 226 AA.
AC Q9QZD2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE COLLAGEN XVIII (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=20227226; PubMed=10766159;
RA Perletti G., Concari P., Giardini R., Marras E., Piccinini F.,
RA Folkman J., Chen L.;
RT "Antitumor activity of endostatin against carcinogen-induced rat
primary mammary tumors";
RL Cancer Res. 60:1793-1796(2000).
DR EMBL; AF189709; AAF00975.1; -.
DR HSSP; P39061; 1KOE.
FT NON_TER 1

SQ SEQUENCE 226 AA; 25350 MW; 3B883C0486C0E949 CRC64;

Query Match 100.0%; Score 45; DB 11; Length 226;
Best Local Similarity 100.0%; Pred. No. 0.096;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||
Db 218 NSFMTSFSK 226

RESULT 4
Q61434 ID Q61434 PRELIMINARY; PRT; 1140 AA.
AC Q61434;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE COLLAGEN (FRAGMENT).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe N., Muragaki Y., Yoshioka H., Inoue H., Ninomiya Y.;
RT "Identification of a novel collagen chain represented by extensive
interruptions in the triple-helical region";
RL Cell. Mol. Biol. Res. 196:576-582(1993).
DR EMBL; D17546; BAA04483.1; -.
DR HSSP; P39061; 1KOE.
DR MGD; MGI:88449; Coll15a1.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 6.
FT NON_TER 1
FT NON_TER 1140
SQ SEQUENCE 1140 AA; 115156 MW; 8B0C7E6862B3BDFF CRC64;

Query Match 100.0%; Score 45; DB 11; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||
Db 1132 NSFMTSFSK 1140

RESULT 5
Q62001 ID Q62001 PRELIMINARY; PRT; 1774 AA.
AC Q62001; Q60672;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE PROCOLLAGEN, TYPE XVIII, ALPHA 1 PRECURSOR (XVIII) COLLAGEN
DE (PROCOLLAGEN, TYPE XVIII, ALPHA 1) (ALPHA-1 TYPE XVIII COLLAGEN).
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PRIMARY TAIL CULTURE;
RX MEDLINE=94245707; PubMed=8188673;
RA Rehn M., Hintikka E., Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
partial structure of the corresponding gene, and comparison of the
alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
chain.";
RL J. Biol. Chem. 269:13929-13935(1994).

RN [2]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=94240112; PubMed=8183894;
RA Rehn M., Pihlajaniemi T.;
RT "Alpha 1(XVII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN [3]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=95181468; PubMed=7876242;
RA Rehn M., Pihlajaniemi T.;
RT "Identification of three N-terminal ends of type XVIII collagen chains
RT and tissue-specific differences in the expression of the corresponding
RT transcripts. The longest form contains a novel motif homologous to rat
RT and Drosophila frizzled proteins.";
RL J. Biol. Chem. 270:4705-4711(1995).
DR EMBL; U03715; AAC52903.1; JOINED.
DR EMBL; U03716; AAC52903.1; JOINED.
DR EMBL; U03718; AAC52903.1; JOINED.
DR EMBL; U34607; AAC52903.1; JOINED.
DR EMBL; U34608; AAC52903.1; JOINED.
DR EMBL; U34609; AAC52903.1; JOINED.
DR EMBL; U34610; AAC52903.1; JOINED.
DR EMBL; U34611; AAC52903.1; JOINED.
DR EMBL; U34612; AAC52903.1; JOINED.
DR EMBL; U34613; AAC52903.1; JOINED.
DR EMBL; U11637; AAC52179.1; -.
DR HSSP; P39061; IKOE.
DR MGD; MGI:88451; Coll18a1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000024; Fz_domain.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 6.
DR Pfam; PF01392; Fz; 1.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00063; FRI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS50038; FZ; 1.
KW Signal.
SQ SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88EF232 CRC64;

Query Match 100.0%; Score 45; DB 11; Length 1774;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 1766 NSFMTSFSK 1774

RESULT 6
Q965U4 PRELIMINARY; PRT; 349 AA.
AC Q965U4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN Y49C4A.5.
GN Y49C4A.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for

investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Johnson D., Strowatt C.;
RT "The sequence of C. elegans cosmid Y49C4A.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024799; AAK72316.1; -.
SQ SEQUENCE 349 AA; 40022 MW; 2D1BB44FCDA01FC1 CRC64;

Query Match 82.2%; Score 37; DB 5; Length 349;
Best Local Similarity 77.8%; Pred. No. 7.2;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 72 SSFLTFSK 80

RESULT 7
Q9XX30 PRELIMINARY; PRT; 128 AA.
AC Q9XX30;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Y39A1A.20 PROTEIN.
GN Y39A1A.20.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wall M.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL031633; CAA21010.1; -.
SQ SEQUENCE 128 AA; 15386 MW; 802CF87A16B72E48 CRC64;

Query Match 75.6%; Score 34; DB 5; Length 128;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 7 NSFLVSFSR 15

RESULT 8
Q9N312 PRELIMINARY; PRT; 263 AA.
AC Q9N312;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 29.6 KDA PROTEIN.
GN Y61A9LA.9.

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Ali J., Dempsey S.;
RT "The sequence of C. elegans cosmid Y61A9LA.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC024843; AAF60836.1; -.
KW Hypothetical protein.
SQ SEQUENCE 263 AA; 29574 MW; E748A1A8FEC70F7C CRC64;

Query Match 75.6%; Score 34; DB 5; Length 263;
Best Local Similarity 77.8%; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 1 NSFMTSFSK 9
|||:||||
DB 221 NSFIFSFSK 229

RESULT 9
O56767 PRELIMINARY; PRT; 296 AA.
AC O56767;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE UL889 PROTEIN (FRAGMENT).
GN UL89.
OS Human cytomegalovirus (strain Towne).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOWNE;
RX MEDLINE=98241707; PubMed=9573236;
RA Krosky P.M., Underwood M.R., Turk S.R., Feng K.W., Jain R.K.,
RA Ptak R.G., Westerman A.C., Biron K.K., Townsend L.B., Drach J.C.;
RT "Resistance of human cytomegalovirus to benzimidazole ribonucleosides
RT maps to two open reading frames: UL89 and UL56.";
RL J. Virol. 72:4721-4728(1998).
DR EMBL; AF047525; AAC40815.1; -.
DR InterPro; IPR003499; DNA_pack_N.
DR Pfam; PF02500; DNA_pack_N; 1.
FT NON_TER 296
SQ SEQUENCE 296 AA; 34293 MW; 3709C6C459206481 CRC64;

Query Match 75.6%; Score 34; DB 12; Length 296;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

* QY 1 NSFMTSFSK 9
|:|||||

Db 128 NFMFTDFKK 136
RESULT 10
Q950Y9 PRELIMINARY; PRT; 604 AA.
AC Q950Y9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYTOCHROME C OXIDASE SUBUNIT 2 (EC 1.9.3.1).
GN COX2.
OS Tetrahymena thermophila.
OG Mitochondrion.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB210;
RA Brunk C.F., Tran A.B., Lee L.C., Li J.;
RT "Complete Sequence of the Mitochondrial Genome of Tetrahymena
RT thermophila and Comparison With the Mitochondrial Genome of
RT Tetrahymena Pyriformis.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF396436; AAK77594.1; -.
KW Oxidoreductase; Mitochondrion.
SQ SEQUENCE 604 AA; 72289 MW; 1CB71D45C50E2F1C CRC64;

Query Match 75.6%; Score 34; DB 8; Length 604;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFS 8
|||:||||
DB 254 NSFITSYS 261

RESULT 11
Q97T55 PRELIMINARY; PRT; 108 AA.
AC Q97T55;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN SP0100.
GN SP0100.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AE007326; AAK74287.1; -.
DR TIGR; SP0100; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 108 AA; 12395 MW; 26BBC0DFB86D1BD9 CRC64;

Query Match 73.3%; Score 33; DB 16; Length 108;
Best Local Similarity 55.6%; Pred. No. 15;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|:|:|:|:
Db 53 NSFLTYSR 61

RESULT 12
Q18743 ID Q18743 PRELIMINARY; PRT; 300 AA.
AC Q18743;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE C50F4.8 PROTEIN.
GN C50F4.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z70750; CRA94743.1; -;
SQ SEQUENCE 300 AA; 33382 MW; FFD1EBBF166C4C48 CRC64;

Query Match 73.3%; Score 33; DB 5; Length 300;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFS 8
|:|:|:|:
Db 73 NDFLTFS 80

RESULT 13
Q9FGD4 ID Q9FGD4 PRELIMINARY; PRT; 792 AA.
AC Q9FGD4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROTEASE-LIKE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026640; BAB08935.1; -;
DR MEROPS; S09.UPA; -;
DR InterPro; IPR000379; Est_lip_thioest_actsite.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002470; Proligo_Prase.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PRINTS; PR00862; PROLIGOPTASE.

SQ SEQUENCE 792 AA; 91001 MW; 960BCA08942F3B51 CRC64;

Query Match 73.3%; Score 33; DB 10; Length 792;
Best Local Similarity 56.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|:|:|:|:
Db 295 NTFSTFSK 303

RESULT 14
Q9AUQ2 ID Q9AUQ2 PRELIMINARY; PRT; 825 AA.
AC Q9AUQ2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 90.7 KDA PROTEIN.
GN OSJNB0033N16.6.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,
RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OSJNB0033N16 genomic sequence.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC082645; AAK18850.1; -;
DR InterPro; IPR002885; PPR.
DR Pfam; PF01535; PPR; 13.
KW Hypothetical protein.
SQ SEQUENCE 825 AA; 90695 MW; 950E6877A55C3A6D CRC64;

Query Match 73.3%; Score 33; DB 10; Length 825;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|:|:|:|:
Db 614 NSLMTSFK 622

RESULT 15
Q9NAE6 ID Q9NAE6 PRELIMINARY; PRT; 183 AA.
AC Q9NAE6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE Y50E8A.12 PROTEIN.
GN Y50E8A.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Steward C.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for

RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL117200; CAB55060.1; --
SQ SEQUENCE 183 AA; 20905 MW; 08064B09B890442F CRC64;

Query Match 71.1%; Score 32; DB 5; Length 183;
Best Local Similarity 55.6%; Pred. No. 42;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSPSK 9
||::|||::
Db 29 NSYVTSFTR 37

Search completed: May 31, 2002, 10:33:08
Job time: 344 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: May 31, 2002, 10:33:26 ; Search time 13.58 Seconds
(without alignments)
25.661 Million cell updates/sec

Title: US-09-589-777C-25

Perfect score: 45

Sequence: 1 NSPMTSFSK 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	45	100.0	1527	1	CALH_MOUSE
2	34	75.6	674	1	VTER_HCMVA
3	32	71.1	573	1	GLB1_MAIZE
4	32	71.1	707	1	HS90_CANAL
5	31	68.9	236	1	MTAP_SULSO
6	31	68.9	273	1	LECG_ARAHY
7	31	68.9	881	1	YFCU_ECOLI
8	30	66.7	243	1	LEC1_ULEEU
9	30	66.7	265	1	LECN_PEA
10	30	66.7	272	1	LCB3_ROBPS
11	30	66.7	285	1	LCB1_ROBPS
12	30	66.7	285	1	LCS1_ROBPS
13	30	66.7	285	1	LCS2_ROBPS
14	30	66.7	286	1	LCB2_ROBPS
15	30	66.7	1002	1	IF2P_YEAST
16	30	66.7	1018	1	DPOG_SCHPO
17	30	66.7	4427	1	PKSL_BACSU
18	29	64.4	168	1	YWY1_CAEEL
19	29	64.4	194	1	PABA_BACSU
20	29	64.4	218	1	Y107_CHLMU
21	29	64.4	229	1	YIEP_ECOLI
22	29	64.4	300	1	T2D3_HAEIN
23	29	64.4	308	1	Y707_HELPY
24	29	64.4	316	1	YCF9_YEAST
25	29	64.4	368	1	COAL_POVLY
26	29	64.4	446	1	YHG4_YEAST
27	29	64.4	456	1	NIFK_METBA
28	29	64.4	492	1	YEOL_YEAST
29	29	64.4	502	1	MTHC_HAEIN
30	29	64.4	518	1	MTHD_HAEIN
31	29	64.4	524	1	TSAG_RICTS
32	29	64.4	529	1	TSAT_RICTS
33	29	64.4	532	1	TSAK_RICTS

34	29	64.4	594	1	RMUC_NEIMA	Q9jwg3 neisseria m
35	29	64.4	700	1	KPC2_DROME	P13677 drosophila
36	29	64.4	779	1	ACON_GRAVE	P49609 gracilaria
37	29	64.4	910	1	HUL5_YEAST	P53119 saccharomyc
38	29	64.4	980	1	SYN_DROME	Q24546 drosophila
39	29	64.4	1351	1	RPOM_YEAST	P13433 saccharomyc
40	29	64.4	1465	1	DPO3_STRPY	Q9fdf9 streptococc
41	29	64.4	4447	1	PKSK_BACSU	P40803 bacillus su
42	29	64.4	4499	1	DYHA_CHLRE	Q39610 chlamydomon
43	28	62.2	185	1	LEC_VICVI	P56625 vicia villo
44	28	62.2	240	1	LECS_VATMA	P81371 vatairea ma
45	28	62.2	249	1	LEC2_ULEEU	P22973 ulex europe

ALIGNMENTS

RESULT 1					
CALH_MOUSE					
ID	CALH_MOUSE	STANDARD;	PRT;	1527	AA.
AC	P39061; Q62002; Q61437;				
DT	01-FEB-1995 (Rel. 31, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].				
GN	COL18A1.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A. (SHORT ISOFORM).				
RC	STRAIN=BALB/C; TISSUE=Liver;				
RX	MEDLINE=94245707; PubMed=8188673;				
RA	Rehn M.V., Hintikka E., Pihlajaniemi T.;				
RT	"Primary structure of the alpha 1 chain of mouse type XVIII collagen,				
RT	partial structure of the corresponding gene, and comparison of the				
RT	alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen				
RT	chain.";				
RL	J. Biol. Chem. 269:13929-13935(1994).				
RN	[2]				
RP	SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).				
RA	Rehn M., Hintikka E., Pihlajaniemi T.;				
RT	"Characterization of the mouse gene for the alpha-1 chain of type				
RT	XVIII collagen (COL18A1) reveals that the three variant N-terminal				
RT	polypeptide forms are transcribed from two widely separated				
RT	promoters.";				
RL	Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).				
RX	MEDLINE=94240112; PubMed=8183894;				
RA	Rehn M.V., Pihlajaniemi T.;				
RT	"Alpha 1(XVIII), a collagen chain with frequent interruptions in the				
RT	collagenous sequence, a distinct tissue distribution, and homology				
RT	with type XV collagen.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).				
RN	[4]				
RP	SEQUENCE OF 240-1527 FROM N.A.				
RC	TISSUE=Liver;				
RX	MEDLINE=94240111; PubMed=8183893;				
RA	Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;				
RT	"Isolation and sequencing of cDNAs for proteins with multiple domains				
RT	of Gly-Xaa-Yaa repeats identify a distinct family of collagenous				
RT	proteins.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).				
RN	[5]				
RP	CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.				
RX	MEDLINE=97160848; PubMed=9008168;				
RA	O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S.,				
RA	Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;				
RT	"Endostatin: an endogenous inhibitor of angiogenesis and tumor				
RT	growth.";				
RL	Cell 88:277-285(1997).				

[6]
RN X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
RX MEDLINE=98169382; PubMed=9501087;
RA Hohenester E., Sasaki T., Olsen B.R., Timpl R.;
RT "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
resolution.";
RL EMBO J. 17:1656-1664(1998).
CC -!- FUNCTION: ENDOSTATIN POTENTLY INHIBITS ENDOTHELIAL CELL
CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCAN INVOLVED IN GROWTH
CC FACTOR SIGNALING.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L16898; AAA37434.1; -.
CC EMBL; U03714; AAA20657.1; -.
CC EMBL; U03715; AAC52901.1; -.
CC EMBL; U34606; AAC52901.1; JOINED.
CC EMBL; U34608; AAC52901.1; JOINED.
CC EMBL; U34609; AAC52901.1; JOINED.
CC EMBL; U34610; AAC52901.1; JOINED.
CC EMBL; U34611; AAC52901.1; JOINED.
CC EMBL; U34612; AAC52901.1; JOINED.
CC EMBL; U34613; AAC52901.1; JOINED.
CC EMBL; U03716; AAC52901.1; JOINED.
CC EMBL; U03718; AAC52901.1; JOINED.
CC EMBL; U03715; AAC52902.1; -.
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; U11636; AAC52178.1; -.
CC EMBL; L22545; AAA19787.1; -.
CC PDB; 1KOE; 16-FEB-99.
CC MGD; MGI:88451; Coll8a1.
CC InterPro; IPR000087; Collagen.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR003129; TSPN.
CC Pfam; PF01391; Collagen; 8.
CC Pfam; PF02210; TSPN; 2.
CC SMART; SM00282; LamG; 1.
CC SMART; SM00210; TSPN; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW 3D-structure.
FT SIGNAL 1 26
FT CHAIN 27 1527
FT CHAIN 1344 1527
FT DOMAIN 27 538
FT DOMAIN 539 565
FT DOMAIN 566 575
FT DOMAIN 576 549
FT DOMAIN 650 673
FT DOMAIN 674 795
FT DOMAIN 796 818
FT DOMAIN 819 901

FT DOMAIN 902 915
FT DOMAIN 916 957
FT DOMAIN 958 970
FT DOMAIN 971 1043
FT DOMAIN 1044 1053
FT DOMAIN 1054 1086
FT DOMAIN 1087 1098
FT DOMAIN 1099 1122
FT DOMAIN 1123 1129
FT DOMAIN 1130 1181
FT DOMAIN 1182 1194
FT DOMAIN 1195 1212
FT DOMAIN 1213 1527
FT CARBOHYD 338 338
FT CARBOHYD 700 700
FT DISULFID 1376 1516
FT DISULFID 1478 1508
FT SITE 1104 1106
FT VARSPLIC 1 212
FT VARSPLIC 213 238
FT CONFLICT 900 900
FT CONFLICT 947 947
FT CONFLICT 964 964
FT CONFLICT 1157 1157
FT CONFLICT 1266 1266
FT CONFLICT 1276 1276
FT CONFLICT 1437 1437
FT SEQUENCE 1527 AA; 156008 MW; 9645045AF140B513 CRC64;
SQ
Query Match 100.0%; Score 45; DB 1; Length 1527;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NSFMSTFSK 9
Db 1519 NSFMSTFSK 1527
RESULT 2
VTER_HCMVA STANDARD; PRT; 674 AA.
AC P16732;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Probable DNA packaging protein.
GN UL89.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,
RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169.";
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL15,
CC HSV-6 ORF12L, EBV UL89, HCMV UL89, EBV BGRF1/BDRF1, AND VZV 42/45.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----


```
RESULT 5
MTAP_SULSO
ID MTAP_SULSO STANDARD; PRT; 236 AA.
AC P50389;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 5'-methylthioadenosine phosphorylase (EC 2.4.2.28) (MTA
DE phosphorylase).
GN MTAP OR SSO2706.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RA Cacciapuoti G.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=95014242; PubMed=7929153;
RA Cacciapuoti G., Porcelli M., Bertoldo C., de Rosa M., Zappia V.;
RT "Purification and characterization of extremely thermophilic and
RT thermostable 5'-methylthioadenosine phosphorylase from the archaeon
RT Sulfolobus solfataricus. Purine nucleoside phosphorylase activity and
RT evidence for intersubunit disulfide bonds."
RL J. Biol. Chem. 269:24762-24769(1994).
RN [3]
RP The complete genome of the crenarchaeon Sulfolobus solfataricus P2."
RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
RN [4]
RP SEQUENCE OF 1-26, AND CHARACTERIZATION.
RX MEDLINE=95014242; PubMed=7929153;
RA Cacciapuoti G., Porcelli M., Bertoldo C., de Rosa M., Zappia V.;
RT "Purification and characterization of extremely thermophilic and
RT thermostable 5'-methylthioadenosine phosphorylase from the archaeon
RT Sulfolobus solfataricus. Purine nucleoside phosphorylase activity and
RT evidence for intersubunit disulfide bonds."
RL J. Biol. Chem. 269:24762-24769(1994).
RN [5]
RP FUNCTION: ENDOWED WITH A BROAD SUBSTRATE SPECIFICITY, BEING ABLE
CC TO PHOSPHOROLYTICALLY CLEAVE INOSINE, GUANOSINE, AND ADENOSINE
CC WITH A BETTER EFFICIENCY THAN MTA. HIGHLY THERMOPHILIC, WITH AN
CC OPTIMUM TEMPERATURE OF 120 DEGREES CELSIUS.
CC CATALYTIC ACTIVITY: 5'-methylthioadenosine + phosphate = adenine +
CC 5-methylthio-D-ribose 1-phosphate.
CC SUBUNIT: HOMOHexamER; DISULFIDE-LINKED.
CC SIMILARITY: BELONGS TO THE PNP/UDP FAMILY 1 OF PHOSPHORYLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

Query Match 68.9%; Score 31; DB 1; Length 236;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY . 1 NSFMTSFSK 9

```
Db 136 NKLVTFSK 144
:|||||
RESULT 6
LECG_ARAHY
ID LECG_ARAHY STANDARD; PRT; 273 AA.
AC P02872;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Galactose-binding lectin precursor (Agglutinin) (PNA).
OS Arachis hypogaea (Peanut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Arachis.
OX NCBI_TaxID=3818;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Seed;
RX MEDLINE=92354708; PubMed=1339358;
RA Rodriguez-Arango E., Arango R., Adar R., Galili G., Sharon N.;
RT "Cloning, sequence analysis and expression in Escherichia coli of the
RT cDNA encoding a precursor of peanut agglutinin."
RL FEBS Lett. 307:185-189(1992).
RN [2]
RP SEQUENCE OF 24-273.
RC TISSUE=Seed;
RX MEDLINE=91192034; PubMed=2013286;
RA Young N.M., Johnston R.A.Z., Watson D.C.;
RT "The amino acid sequence of peanut agglutinin."
RL Eur. J. Biochem. 196:631-637(1991).
RN [3]
RP PRELIMINARY PARTIAL SEQUENCE.
RC STRAIN=CV. SHULAMIT;
RA Lauwereys M., Foriers A., Sharon N., Strosberg A.D.;
RT "Sequence studies of peanut agglutinin."
RL FEBS Lett. 181:241-244(1985).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS).
RX MEDLINE=96228253; PubMed=8656429;
RA Banerjee R., Das K., Ravishankar R., Suguna K., Surolia A.,
RA Vijayan M.;
RT "Conformation, protein-carbohydrate interactions and a novel subunit
RT association in the refined structure of peanut lectin-lactose
RT complex."
RL J. Mol. Biol. 259:281-296(1996).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX MEDLINE=99348591; PubMed=10417405;
RA Ravishankar R., Suguna K., Surolia A., Vijayan M.;
RT "Structures of the complexes of peanut lectin with methyl-beta-
RT galactose and N-acetyllactosamine and a comparative study of
RT carbohydrate binding in Gal/GalNAc-specific legume lectins."
RL Acta Crystallogr. D 55:1375-1382(1999).
CC -!- FUNCTION: D-GALACTOSE SPECIFIC LECTIN.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- MISCELLANEOUS: BINDS ONE MANGANESE (OR OTHER TRANSITION METAL) ION
CC AND ONE CALCIUM ION. THE METAL IONS ARE ESSENTIAL FOR THE
CC SACCHARIDE-BINDING AND CELL-AGGLUTINATING ACTIVITIES.
CC -!- SIMILARITY: BELONGS TO THE LEGUMINOUS LECTIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR PIR; S14765; S14765.
DR PDB; 2PEL; 08-DEC-96.
DR PDB; 2TEP; 08-APR-99.
DR PDB; 1BZW; 11-NOV-98.
DR InterPro; IPR000985; Lectin_lega.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF00138; lectin_legA; 1.
DR Pfam; PF00139; lectin_legB; 1.
DR ProDom; PD000671; Lectin_legA; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; 1.
DR PROSITE; PS00308; LECTIN_LEGUME_ALPHA; FALSE_NEG.
KW Lectin; Calcium; Manganese; 3D-structure; Signal.
FT SIGNAL 1 23
FT CHAIN 24 273 GALACTOSE-BINDING LECTIN.
FT METAL 144 144 MANGANESE.
FT METAL 146 146 MANGANESE AND CALCIUM.
FT METAL 148 148 CALCIUM.
FT METAL 150 150 CALCIUM.
FT METAL 155 155 MANGANESE AND CALCIUM.
FT METAL 160 160 MANGANESE.
FT VARIANT 115 115 E -> V (IN MINOR FORM).
FT VARIANT 172 172 K -> A (IN MINOR FORM).
FT VARIANT 185 185 K -> I (IN MINOR FORM).
FT VARIANT 235 236 LG -> RA (IN MINOR FORM).
FT SEQUENCE 273 AA; 29325 MW; 05A0B1A8FAC7B159 CRC64;

Query Match 68.9%; Score 31; DB 1; Length 273;
Best Local Similarity 85.7%; Pred. No. 26;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFMTSFS 8
Db 87 SFLTSFS 93
||| |||

RESULT 7
YFCU_ECOLI
ID YFCU_ECOLI STANDARD; PRT; 881 AA.
AC P77196; P77532;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane usher protein yfcU precursor.
GN YFCU OR B2337/B2338.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Sato Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horluchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
```

```
CC -!- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF A FIMBRIAL
CC SUBUNIT ACROSS THE OUTER MEMBRANE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION, AN IN-FRAME STOP CODON
CC WAS READ OVER IN POSITION 579.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000322; AAC75398.1; ALT_TERM.
CC EMBL; AE000322; AAC75397.1; ALT_INIT.
CC EMBL; D90864; BAA16191.1; ALT_INIT.
CC EMBL; D90864; BAA16192.1; ALT_TERM.
CC EcoGene; EG14124; yfcU.
CC InterPro; IPR000015; Fimb_usher.
CC Pfam; PF00577; Usher; 1.
CC PROSITE; PS01151; FIMBRIAL_USHER; 1.
KW Hypothetical protein; Outer membrane; Transmembrane; Fimbria;
KW Transport; Signal; Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 881 HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN
FT YFCU.
FT SEQUENCE 881 AA; 97437 MW; 23DE53B67541E041 CRC64;

Query Match 68.9%; Score 31; DB 1; Length 881;
Best Local Similarity 66.7%; Pred. No. 90;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 457 NSFRVSYSK 465
||| |||

RESULT 8
LEC1_ULEEU
ID LEC1_ULEEU STANDARD; PRT; 243 AA.
AC P22972;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Anti-H(O) lectin I (UEA-I).
OS Ulex europeus (Furze).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Ulex.
OX NCBI_TaxID=3902;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RX MEDLINE=91331980; PubMed=1869520;
RA Konami Y., Yamamoto K., Osawa T.;
RT "The primary structures of two types of the Ulex europeus seed
RT lectin.";
RL J. Biochem. 109:650-658(1991).
RN [2]
RP SEQUENCE OF 1-34.
RX MEDLINE=91315756; PubMed=1859632;
RA Konami Y., Yamamoto K., Osawa T.;
RT "Purification and characterization of a new type lactose-binding Ulex
RT europeus lectin by affinity chromatography.";
RL Biol. Chem. Hoppe-Seyler 372:95-102(1991).
CC -!- FUNCTION: L-FUCOSE SPECIFIC LECTIN.
CC -!- SIMILARITY: BELONGS TO THE LEGUMINOUS LECTIN FAMILY.
CC PIR; JX0162; JX0162.
DR PIR; S13437; S13437.
```

```
DR HSSP; P12306; ILQE.
DR InterPro; IPR000985; Lectin_legA.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF00138; lectin_legA; 1.
DR Pfam; PF00139; lectin_legB; 1.
DR ProDom; PD000671; Lectin_legA; 1.
DR ProSITE; PS00307; LECTIN_LEGUME_BETA; 1.
DR ProSITE; PS00308; LECTIN_LEGUME_ALPHA; 1.
KW Lectin; Calcium; Manganese; Glycoprotein.
FT METAL 126 126 MANGANESE (BY SIMILARITY).
FT METAL 128 128 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 135 135 CALCIUM (BY SIMILARITY).
FT METAL 138 138 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 143 143 MANGANESE (BY SIMILARITY).
FT CARBOHYD 10 10 N-LINKED (GLCNAC. . .) (PARTIAL).
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .).
SQ SEQUENCE 243 AA; 26669 MW; 74F2D74A8A2CF8E1 CRC64;

Query Match 66.7%; Score 30; DB 1; Length 243;
Best Local Similarity 85.7%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFMTSFS 8
Db 67 SFITSFS 73

RESULT 9
LECN_PEA
ID LECN_PEA STANDARD; PRT; 265 AA.
AC P16270;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Nonseed lectin precursor.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, ALASKA;
RX MEDLINE=92256807; PubMed=1581566;
RA Pak J.H., Hendrickson T., Dobres M.S.;
RT "A developmentally regulated bud specific transcript in Pisum
sativum.";
RL Plant Physiol. 89:833-838(1989).
RN [2]
RP SEQUENCE OF 131-202 FROM N.A.
RC STRAIN=CV, ALASKA;
RA Dobres M.S., Thompson W.F.;
RT "A developmentally regulated bud specific transcript in pea has
sequence similarity to seed lectins.";
RL Plant Physiol. 89:833-838(1989).
CC -!- SUBUNIT: MONOMER.
CC -!- TISSUE SPECIFICITY: MOST HIGHLY EXPRESSED IN THE EPIDERMAL
CC LAYER OF DEVELOPING SHOOT TIPS.
CC -!- SIMILARITY: BELONGS TO THE LEGUMINOUS LECTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L11745; AAA33675.1; -.
CC EMBL; M25072; AAA33681.1; -.
CC PIR; A44975; A44975.
CC HSSP; P02872; 2PEL.
CC -----
```

```
DR InterPro; IPR000985; Lectin_legA.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF00138; lectin_legA; 1.
DR Pfam; PF00139; lectin_legB; 1.
DR ProDom; PD000671; Lectin_legA; 1.
DR ProSITE; PS00307; LECTIN_LEGUME_BETA; FALSE_NEG.
DR ProSITE; PS00308; LECTIN_LEGUME_ALPHA; FALSE_NEG.
KW Lectin; Signal; Multigene family; Manganese; Calcium; Glycoprotein.
FT SIGNAL 1 21 OR 23 (POTENTIAL).
FT CHAIN 22 265 NONSEED LECTIN.
FT CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 265 AA; 28530 MW; BC3CC98D7DBCA8B2 CRC64;

Query Match 66.7%; Score 30; DB 1; Length 265;
Best Local Similarity 85.7%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFMTSFS 8
Db 93 SFVTFSFS 99

RESULT 10
LCB3_ROBPS
ID LCB3_ROBPS STANDARD; PRT; 272 AA.
AC Q41160;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Putative bark agglutinin LECRPA3 precursor (Fragment).
OS Robinia pseudoacacia (Black locust).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Robinieae; Robinia.
OX NCBI_TaxID=35938;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bark;
RX MEDLINE=95232198; PubMed=7716244;
RA van Damme E.J.M., Barre A., Smeets K., Torrekens S., van Leuven F.,
RA Rouge P., Peumans W.J.;
RT "The bark of Robinia pseudoacacia contains a complex mixture of
lectins. Characterization of the proteins and the cDNA clones.";
RL Plant Physiol. 107:833-843(1995).
CC -!- FUNCTION: BARK LECTINS ARE STORAGE PROTEIN THAT PROBABLY MAINTAINS
CC STOCKS OF NITROGEN DURING DORMANT PERIOD. SELF-AGGREGATABLE
CC MOLECULES THAT CAN BIND THEIR OWN CARBOHYDRATE SIDE CHAINS. THEY
CC COULD ALSO PLAY A ROLE IN THE PLANT'S DEFENCE AGAINST PHYTOPHAGOUS
CC INVERTEBRATES OR HERBIVOROUS HIGHER ANIMALS.
CC -!- SUBUNIT: HOMOTETRAMER (POTENTIAL).
CC -!- TISSUE SPECIFICITY: WEAK EXPRESSION IN BARK. THE LECTIN
CC ACCUMULATES IN THE INNER BARK IN AUTUMN.
CC -!- SIMILARITY: TO OTHER LEGUMINOUS LECTINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U12784; AAA80183.1; ALT_INIT.
CC HSSP; P02872; 2PEL.
DR InterPro; IPR000985; Lectin_legA.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF00138; lectin_legA; 1.
DR Pfam; PF00139; lectin_legB; 1.
DR ProDom; PD000671; Lectin_legA; 1.
DR ProSITE; PS00307; LECTIN_LEGUME_BETA; 1.
DR ProSITE; PS00308; LECTIN_LEGUME_ALPHA; 1.
DR HSSP; P02872; 2PEL.
```

KW Lectin; Calcium; Manganese; Glycoprotein; Signal; Multigene family.
FT NON_TER 1 1
FT SIGNAL <1 29
FT CHAIN 30 272 PUTATIVE BARK AGGLUTININ LECRPA3.
FT METAL 150 150 MANGANESE (BY SIMILARITY).
FT METAL 152 152 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 154 154 CALCIUM (BY SIMILARITY).
FT METAL 156 156 CALCIUM (BY SIMILARITY).
FT METAL 159 159 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 164 164 MANGANESE (BY SIMILARITY).
FT CARBOHYD 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 272 AA; 29195 MW; BE8B2834EA3100D4 CRC64;

Query Match 66.7%; Score 30; DB 1; Length 272;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFMTSFS 8
II:IIII
Db 94 SFVTSFS 100

RESULT 11
LCBI_ROBPS
ID LCBI_ROBPS STANDARD; PRT; 285 AA.
AC Q41159;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bark agglutinin I, polypeptide A precursor (RPBAI) (LECRPA1).
OS Robinia pseudoacacia (Black locust).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Robinieae; Robinia.
OX NCBI_TaxID=35938;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-50.
RC TISSUE=Bark;
RX MEDLINE=95232198; PubMed=7716244;
RA van Damme E.J.M., Barre A., Smeets K., Torrekens S., van Leuven F.,
RA Rouge P., Peumans W.J.;
RT "The bark of Robinia pseudoacacia contains a complex mixture of
lectins. Characterization of the proteins and the cDNA clones."
RL Plant Physiol. 107:833-843(1995).
CC -!- FUNCTION: N-ACETYL-D-GALACTOSAMINE SPECIFIC LECTIN.
CC -!- FUNCTION: BARK LECTINS ARE STORAGE PROTEIN THAT PROBABLY MAINTAINS
CC STOCKS OF NITROGEN DURING DORMANT PERIOD. SELF-AGGREGATABLE
CC MOLECULES THAT CAN BIND THEIR OWN CARBOHYDRATE SIDE CHAINS. THEY
CC COULD ALSO PLAY A ROLE IN THE PLANT'S DEFENSE AGAINST PHYTOPHAGOUS
CC INVERTEBRATES OR HERBIVOROUS HIGHER ANIMALS.
CC -!- SUBUNIT: RPBAI IS COMPOSED OF TWO POLYPEPTIDES, A AND B, THAT
CC ASSOCIATE INTO FIVE DIFFERENT TETRAMERIC ISOLECTINS. THE A4
CC COMBINATION IS THE ONLY ONE DEVOID OF AGGLUTINATION ACTIVITY.
CC ISOFORM B4 DISPLAYS MAXIMAL AGGLUTINATION ACTIVITY.
CC -!- TISSUE SPECIFICITY: STRONG EXPRESSION IN SEED. LOWER LEVELS IN THE
CC FLOWER, AND THE BARK OF THE ROOTS. NO EXPRESSION IN LEAF. THE
CC LECTIN ACCUMULATES IN THE INNER BARK IN AUTUMN.
CC -!- SIMILARITY: TO OTHER LEGUMINOUS LECTINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U12782; AAA80181.1; -.
DR HSSP; P19588; 1LUL.
DR InterPro; IPR000985; Lectin_legA.

DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF00138; lectin_legA; 1.
DR Pfam; PF00139; lectin_legB; 1.
DR ProDom; PD000671; Lectin_legA; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; 1.
DR PROSITE; PS00308; LECTIN_LEGUME_ALPHA; 1.
KW Lectin; Calcium; Manganese; Glycoprotein; Signal; Multigene family.
FT SIGNAL 1 31
FT CHAIN 32 285 BARK AGGLUTININ I, POLYPEPTIDE A.
FT METAL 156 156 MANGANESE (BY SIMILARITY).
FT METAL 158 158 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 160 160 CALCIUM (BY SIMILARITY).
FT METAL 162 162 CALCIUM (BY SIMILARITY).
FT METAL 166 166 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 171 171 MANGANESE (BY SIMILARITY).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 285 AA; 30928 MW; 49382E50EEF27282 CRC64;

Query Match 66.7%; Score 30; DB 1; Length 285;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFMTSFS 8
II:IIII
Db 99 SFVTSFS 105

RESULT 12
LCBI_ROBPS
ID LCBI_ROBPS STANDARD; PRT; 285 AA.
AC Q41162;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Seed agglutinin I precursor (RPSAI) (LECRPAS1).
OS Robinia pseudoacacia (Black locust).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Robinieae; Robinia.
OX NCBI_TaxID=35938;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-51.
RC TISSUE=Seed;
RX MEDLINE=96191285; PubMed=8616218;
RA van Damme E.J.M., Barre A., Rouge P., van Leuven F., Peumans W.J.;
RT "The seed lectins of black locust (Robinia pseudoacacia) are encoded
by two genes which differ from the bark lectin genes."
RL Plant Mol. Biol. 29:1197-1210(1995).
CC -!- FUNCTION: SEED LECTIN.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN SEED.
CC -!- PTM: THREE OF THE FOUR POTENTIAL GLYCOSYLATION SITES ARE OCCUPIED.
CC -!- SIMILARITY: TO OTHER LEGUMINOUS LECTINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U24250; AAC49272.1; -.
DR HSSP; P19588; 1LUL.
DR InterPro; IPR000985; Lectin_legA.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF00138; lectin_legA; 1.
DR Pfam; PF00139; lectin_legB; 1.
DR ProDom; PD000671; Lectin_legA; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; 1.
DR PROSITE; PS00308; LECTIN_LEGUME_ALPHA; 1.

KW Lectin; Calcium; Manganese; Glycoprotein; Signal; Multigene family.
FT SIGNAL 1 31
FT CHAIN 32 285 SEED AGGLUTININ I.
FT METAL 156 156 MANGANESE (BY SIMILARITY).
FT METAL 158 158 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 162 162 CALCIUM (BY SIMILARITY).
FT METAL 166 166 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 171 171 MANGANESE (BY SIMILARITY).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 285 AA; 30943 MW; 6AE82CDC920224CE CRC64;

Query Match 66.7%; Score 30; DB 1; Length 285;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFMTSFS 8
Db 99 SFVTSFS 105
||:||||

RESULT 13
LCS2_ROBPS
ID LCS2_ROBPS STANDARD; PRT; 285 AA.
AC Q41161;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Seed agglutinin II precursor (RPSAII) (LECRPAS2).
OS Robinia pseudoacacia (Black locust).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Robinieae; Robinia.
OX NCBI_TaxID=35938;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-50.
RC TISSUE=Seed;
RX MEDLINE=96191285; PubMed=8616218;
RA van Damme E.J.M., Barre A., Rouge P., van Leuven F., Peumans W.J.;
RT "The seed lectins of black locust (Robinia pseudoacacia) are encoded
by two genes which differ from the bark lectin genes.";
RL Plant Mol. Biol. 29:1197-1210(1995).
CC -!- FUNCTION: SEED LECTIN.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- TISSUE SPECIFICITY: EXPRESSION IN SEED.
CC -!- PTM: MOSTLY FOUND IN NON-GLYCOSYLATED FORM.
CC -!- SIMILARITY: TO OTHER LEGUMINOUS LECTINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch>)
or send an email to license@isb-sib.ch.

DR EMBL; U24249; AAC49271.1; -.
DR HSSP; P19588; 1LJUL.
DR InterPro; IPR000985; Lectin_legA.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF00138; lectin_legA; 1.
DR Pfam; PF00139; lectin_legB; 1.
DR ProDom; PD000671; Lectin_legA; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; 1.
DR PROSITE; PS00308; LECTIN_LEGUME_ALPHA; 1.
KW Lectin; Calcium; Manganese; Glycoprotein; Signal; Multigene family.
FT SIGNAL 1 31 SEED AGGLUTININ II.
FT CHAIN 32 285
FT METAL 156 156 MANGANESE (BY SIMILARITY).
FT METAL 158 158 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 162 162 CALCIUM (BY SIMILARITY).
FT METAL 166 166 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 171 171 MANGANESE (BY SIMILARITY).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 285 AA; 30943 MW; 6AE82CDC920224CE CRC64;

Query Match 66.7%; Score 30; DB 1; Length 285;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFMTSFS 8
Db 99 SFVTSFS 105
||:||||

RESULT 14
LCS2_ROBPS
ID LCS2_ROBPS STANDARD; PRT; 286 AA.
AC Q42372;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bark agglutinin I, polypeptide B precursor (RPBAI) (LECRPA2).
OS Robinia pseudoacacia (Black locust).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Robinieae; Robinia.
OX NCBI_TaxID=35938;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-49.
RC TISSUE=Bark;
RX MEDLINE=95232198; PubMed=7716244;
RA van Damme E.J.M., Barre A., Smeets K., Torrekens S., van Leuven F.,
Rouge P., Peumans W.J.;
RT "The bark of Robinia pseudoacacia contains a complex mixture of
lectins. Characterization of the proteins and the cDNA clones.";
RL Plant Mol. Biol. 25:845-853(1994).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-49.
RX MEDLINE=95232198; PubMed=7716244;
RA van Damme E.J.M., Barre A., Smeets K., Torrekens S., van Leuven F.,
Rouge P., Peumans W.J.;
RT "The bark of Robinia pseudoacacia contains a complex mixture of
lectins. Characterization of the proteins and the cDNA clones.";
RL Plant Physiol. 107:833-843(1995).
RN [3]
RP SEQUENCE OF 32-51.
RC TISSUE=Bark;
RA Tazaki K., Yoshida K.;
RT "The bark lectin of Robinia pseudoacacia: purification and partial
characterization.";
RL Plant Cell Physiol. 33:125-129(1992).
CC -!- FUNCTION: BARK LECTINS ARE STORAGE PROTEIN THAT PROBABLY MAINTAINS
STOCKS OF NITROGEN DURING DORMANT PERIOD. SELF-AGGREGATABLE
MOLECULES THAT CAN BIND THEIR OWN CARBOHYDRATE SIDE CHAINS. THEY
COULD ALSO PLAY A ROLE IN THE PLANT'S DEFENCE AGAINST PHYTOPHAGOUS
INVERTEBRATES OR HERBIVOROUS HIGHER ANIMALS.
CC -!- SUBUNIT: RPBAI IS COMPOSED OF TWO POLYPEPTIDES, A AND B, THAT
ASSOCIATE INTO FIVE DIFFERENT TETRAMERIC ISOLECTINS. THE A4
COMBINATION IS THE ONLY ONE DEVOID OF AGGLUTINATION ACTIVITY.
CC ISOFORM B4 DISPLAYS MAXIMAL AGGLUTINATION ACTIVITY.
CC -!- TISSUE SPECIFICITY: MOSTLY IN THE AXIAL AND RAY PARENCHYMAL CELLS
OF THE INNER BARK. FEWER IN THE AXIAL AND RAY PARENCHYMAL CELLS OF
THE XYLEM. STRONG EXPRESSION IN BARK. THE LECTIN ACCUMULATES IN
THE INNER BARK IN AUTUMN AND WINTER AND DISAPPEARS IN MAY.
CC -!- SIMILARITY: TO OTHER LEGUMINOUS LECTINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch>)
or send an email to license@isb-sib.ch.

FT METAL 162 162 CALCIUM (BY SIMILARITY).
FT METAL 166 166 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 171 171 MANGANESE (BY SIMILARITY).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 285 AA; 31021 MW; 2C0B3249620294DE CRC64;

Query Match 66.7%; Score 30; DB 1; Length 285;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFMTSFS 8
Db 99 SFVTSFS 105
||:||||

RESULT 14
LCS2_ROBPS
ID LCS2_ROBPS STANDARD; PRT; 286 AA.
AC Q42372;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bark agglutinin I, polypeptide B precursor (RPBAI) (LECRPA2).
OS Robinia pseudoacacia (Black locust).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Robinieae; Robinia.
OX NCBI_TaxID=35938;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bark;
RX MEDLINE=94355657; PubMed=7915553;
RA Yoshida K., Baba K., Yamamoto N., Tazaki K.;
RT "Cloning of a lectin cDNA and seasonal changes in levels of the lectin
and its mRNA in the inner bark of Robinia pseudoacacia.";
RL Plant Mol. Biol. 25:845-853(1994).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-49.
RX MEDLINE=95232198; PubMed=7716244;
RA van Damme E.J.M., Barre A., Smeets K., Torrekens S., van Leuven F.,
Rouge P., Peumans W.J.;
RT "The bark of Robinia pseudoacacia contains a complex mixture of
lectins. Characterization of the proteins and the cDNA clones.";
RL Plant Physiol. 107:833-843(1995).
RN [3]
RP SEQUENCE OF 32-51.
RC TISSUE=Bark;
RA Tazaki K., Yoshida K.;
RT "The bark lectin of Robinia pseudoacacia: purification and partial
characterization.";
RL Plant Cell Physiol. 33:125-129(1992).
CC -!- FUNCTION: BARK LECTINS ARE STORAGE PROTEIN THAT PROBABLY MAINTAINS
STOCKS OF NITROGEN DURING DORMANT PERIOD. SELF-AGGREGATABLE
MOLECULES THAT CAN BIND THEIR OWN CARBOHYDRATE SIDE CHAINS. THEY
COULD ALSO PLAY A ROLE IN THE PLANT'S DEFENCE AGAINST PHYTOPHAGOUS
INVERTEBRATES OR HERBIVOROUS HIGHER ANIMALS.
CC -!- SUBUNIT: RPBAI IS COMPOSED OF TWO POLYPEPTIDES, A AND B, THAT
ASSOCIATE INTO FIVE DIFFERENT TETRAMERIC ISOLECTINS. THE A4
COMBINATION IS THE ONLY ONE DEVOID OF AGGLUTINATION ACTIVITY.
CC ISOFORM B4 DISPLAYS MAXIMAL AGGLUTINATION ACTIVITY.
CC -!- TISSUE SPECIFICITY: MOSTLY IN THE AXIAL AND RAY PARENCHYMAL CELLS
OF THE INNER BARK. FEWER IN THE AXIAL AND RAY PARENCHYMAL CELLS OF
THE XYLEM. STRONG EXPRESSION IN BARK. THE LECTIN ACCUMULATES IN
THE INNER BARK IN AUTUMN AND WINTER AND DISAPPEARS IN MAY.
CC -!- SIMILARITY: TO OTHER LEGUMINOUS LECTINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch>)
or send an email to license@isb-sib.ch.

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D17757; BAA04604.1; -.
DR EMBL; U12783; AAA80182.1; -.
DR HSSP; P19588; 1LJL.
DR InterPro; IPR000985; Lectin_legA.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF00138; lectin_legA; 1.
DR Pfam; PF00139; lectin_legB; 1.
DR ProDom; PD000671; Lectin_legA; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; 1.
DR PROSITE; PS00308; LECTIN_LEGUME_ALPHA; 1.
KW Lectin; Calcium; Manganese; Glycoprotein; Signal; Multigene family.
FT SIGNAL 1 31
FT CHAIN 32 286 BARK AGGLUTININ I, POLYPEPTIDE B.
FT METAL 157 157 MANGANESE (BY SIMILARITY).
FT METAL 159 159 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 163 163 CALCIUM (BY SIMILARITY).
FT METAL 167 167 MANGANESE AND CALCIUM (BY SIMILARITY).
FT METAL 172 172 MANGANESE (BY SIMILARITY).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 43 46 KHSQ -> MPNE (IN REF. 2; AA SEQUENCE).
FT CONFLICT 48 48 D -> W (IN REF. 3).
SQ SEQUENCE 286 AA; 31211 MW; 467E37661D1DC1E6 CRC64;

Query Match 66.7%; Score 30; DB 1; Length 286;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFMTSFS 8
Db 99 SFVTSFS 105

RESULT 15
IF2P_YEAST
ID IF2P_YEAST STANDARD; PRT; 1002 AA.
AC P39730;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Translation initiation factor IF-2.
GN YAL035W OR FUN12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94357438; PubMed=8076820;
RA Suttrave P., Shafer B.K., Strathern J.N., Hughes S.H.;
RT "Isolation, identification and characterization of the FUN12 gene of
RT Saccharomyces cerevisiae.";
RL Gene 146:209-213(1994).
RN [3]
RP REVISIONS.
RC STRAIN=S288C / AB972;
RA Vo D.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=98288357; PubMed=9624054;
```

```
RA Choi S.K., Lee J.H., Zoll W.L., Merrick W.C., Dever T.E.;
RT "Promotion of met-tRNAMet binding to ribosomes by YIF2, a bacterial
RT IF2 homolog in yeast.";
RL Science 280:1757-1760(1998).
CC -!- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING
CC THE BINDING OF THE FORMYL METHIONINE-tRNA TO RIBOSOMES. SEEMS TO
CC FUNCTION ALONG WITH EIF-2.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U12980; AAC04996.1; -.
CC EMBL; L29389; AAA57228.1; ALT_SEQ.
CC SGD; S0000033; FUN12.
CC InterPro; IPR000795; GTP_EFTU.
CC InterPro; IPR004161; GTP_EFTU_D2.
CC InterPro; IPR000178; IF2.
CC Pfam; PF00009; GTP_EFTU; 1.
CC Pfam; PF03144; GTP_EFTU_D2; 2.
CC PROSITE; PS01176; IF2; FALSE_NEG.
KW Initiation factor; Protein biosynthesis; GTP-binding.
FT DOMAIN 361 371 POLY-GLU.
FT NP_BIND 412 419 GTP (BY SIMILARITY).
SQ SEQUENCE 1002 AA; 112268 MW; 1A496195DAE1C283 CRC64;

Query Match 66.7%; Score 30; DB 1; Length 1002;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 544 NSFRDSFAK 552

Search completed: May 31, 2002, 10:33:28
Job time: 334 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:28:24 ; Search time 26.62 Seconds
(without alignments)
32.487 Million cell updates/sec

Title: US-09-589-777C-25
Perfect score: 45
Sequence: 1 NSFMTSFSK 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	1315	2 A56101	collagen alpha 1(X
2	45	100.0	1774	2 B56101	collagen alpha 1(X
3	34	75.6	128	2 T26721	hypothetical prote
4	34	75.6	724	1 QQBEI5	UL89 protein - hum
5	33	73.3	108	2 F95Q11	conserved hypothet
6	33	73.3	108	2 A97883	conserved hypothet
7	33	73.3	300	2 T20115	hypothetical prote
8	32	71.1	171	2 JQ0699	hypothetical 19K p
9	32	71.1	183	2 T31621	hypothetical prote
10	32	71.1	203	2 H97015	hypothetical prote
11	32	71.1	395	2 A61073	glutamine amidotra
12	32	71.1	573	2 A53234	heat shock protein
13	32	71.1	582	2 B53234	globulin-1S, GLB1S
14	32	71.1	655	2 H96692	vicilin-like stora
15	32	71.1	1056	2 T28636	probable receptor
16	32	71.1	2161	1 A45389	zinc-metallopeptid
17	32	71.1	7829	2 T15789	genome polypeptide
18	31	68.9	236	2 S58291	hypothetical prote
19	31	68.9	265	2 C86851	5'-methylthioaden
20	31	68.9	273	2 S24044	hypothetical prote
21	31	68.9	312	2 T00160	lectin precursor -
22	31	68.9	312	2 S32211	leucocidin chain S
23	31	68.9	368	2 D95413	conserved hypothet
24	31	68.9	578	2 H65006	hypothetical prote
25	31	68.9	585	2 E97232	probable signal tr
26	31	68.9	598	2 H69011	hypothetical prote
27	31	68.9	720	2 G97229	pullulanase [impor
28	31	68.9	814	2 T00740	hypothetical prote
29	31	68.9	836	2 T21631	hypothetical prote

30	30	66.7	108	2 C81299	hypothetical prote
31	30	66.7	156	2 T28997	hypothetical prote
32	30	66.7	243	2 JX0162	lectin I - furze
33	30	66.7	265	2 S20988	lectin - garden pe
34	30	66.7	270	2 T06528	lectin - garden pe
35	30	66.7	285	2 S62690	agglutinin I precu
36	30	66.7	285	2 S62691	agglutinin II prec
37	30	66.7	286	2 S48033	lectin precursor -
38	30	66.7	290	2 T18795	hypothetical prote
39	30	66.7	296	2 B90270	sugar ABC transpor
40	30	66.7	302	2 AF1021	probable membrane-
41	30	66.7	308	2 T29756	hypothetical prote
42	30	66.7	339	2 G97199	probable permease
43	30	66.7	344	2 D86362	hypothetical prote
44	30	66.7	348	2 F89765	hypothetical prote
45	30	66.7	368	2 A69774	integrase homolog

ALIGNMENTS

RESULT 1
A56101
collagen alpha 1(XVIII) chain precursor, short splice form - mouse
N:Contains: endostatin
C:Species: Mus musculus (house mouse)
C:Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
C:Accession: A56101; A58371; S72450; S65595
R:Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A:Title: Identification of three N-terminal ends of type XVIII collagen chains and t:
tif homologous to rat and Drosophila frizzled proteins.
A:Reference number: A56101; MUID:95181468
A:Accession: A56101
A:Molecule type: mRNA
A:Residues: 1-103 <REH1>
A:Cross-references: GB:U11636; NID:G618427; PIDN:AAC52178.1; PID:G618428
R:Rehn, M.; Pihlajaniemi, T.
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994
A:Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagen
A:Reference number: A58371; MUID:94240112
A:Accession: A58371
A:Molecule type: mRNA
A:Residues: 1-928 <REH2>
A:Cross-references: GB:L16898; NID:G404754; PIDN:AAA37434.1; PID:G553894
R:Oh, S.P.; Kanagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A:Reference number: S72450
A:Accession: S72450
A:Molecule type: mRNA
A:Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHW>
A:Cross-references: EMBL:L22545; NID:G348968; PIDN:AAAL9787.1; PID:G511298
R:Oh, S.P.; Kanagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-
A:Reference number: A58370; MUID:94240111
A:Accession: S65595
A:Molecule type: mRNA
A:Residues: 28-1315 <OHS>
A:Cross-references: EMBL:L22545
C:Comment: Prolines and lysines at the third position of the tripeptide repeating uni-
lated and subsequently O-glycosylated.
C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in f
C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of
ay be useful in treating solid tumors.
C:Genetics:
A:Gene: MGI:Coll8a1
A:Cross-references: MGI:711175
A:Map position: 10:41.0
C:Superfamily: unassigned collagens
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog
F:1-25/Domain: signal sequence #status predicted <SIG>
F:24-235/Region: thrombospondin amino-terminal similarity

F;26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M
F;327-353/Domain: collagenous #status predicted <CO1>
F;364-437/Domain: collagenous #status predicted <CO2>
F;462-583/Domain: collagenous #status predicted <CO3>
F;607-689/Domain: collagenous #status predicted <CO4>
F;704-745/Domain: collagenous #status predicted <CO5>
F;759-831/Domain: collagenous #status predicted <CO6>
F;842-874/Domain: collagenous #status predicted <CO7>
F;887-910/Domain: collagenous #status predicted <CO8>
F;892-894/Region: cell attachment (R-G-D) motif
F;918-969/Domain: collagenous #status predicted <CO9>
F;983-1000/Domain: collagenous #status predicted <CO10>
F;1132-1315/Product: endostatin #status predicted <EST>
F;1139-1315/Region: multiplexin collagen carboxyl-terminal similarity
F;126-488/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;172-228/Disulfide bonds: #status predicted
F;240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 45; DB 2; Length 1315;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||||

Db 1307 NSFMTSFSK 1315

RESULT 2
B56101
collagen alpha 1(XVIII) chain precursor, long splice form - mouse
N:Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
C:Species: Mus musculus (house mouse)
C:Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
C:Accession: B56101; C56101; S72450; S65595; PN0675; A54072; A58816
R;Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995
A:Title: Identification of three N-terminal ends of type XVIII collagen chains and tissue homologs to rat and Drosophila frizzled proteins.
A:Reference number: A56101; MUID:95181468
A:Accession: B56101
A:Molecule type: mRNA
A:Residues: 1-562 <REH1>
A:Cross-references: GB:U11637; NID:g618429; PIDN:AAC52179.1; PID:g618430
A:Experimental source: splice form clone PE17.24
A:Accession: C56101
A:Molecule type: mRNA
A:Residues: 1-239,487-562 <REH2>
A:Cross-references: GB:U11637; NID:g618429
A:Experimental source: splice form clones PE8.1, PE19, PE15.2
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993
A:Reference number: S72450
A:Accession: S72450
A:Molecule type: mRNA
A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-16
A:Cross-references: EMBL:L22545; NID:g348968; PIDN:AAA19787.1; PID:g511298
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A:Reference number: A58370; MUID:94240111
A:Accession: S65595
A:Molecule type: mRNA
A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>
A:Cross-references: EMBL:L22545
R;Abe, N.; Muragaki, Y.; Yoshioaka, H.; Inoue, H.; Ninomiya, Y.
Biochem. Biophys. Res. Commun. 196, 576-582, 1993
A:Title: Identification of a novel collagen chain represented by extensive interruptions
A:Reference number: PN0675; MUID:94059075
A:Accession: PN0675
A:Molecule type: mRNA
A:Residues: 635-1774 <ABE>

R;Rehn, M.; Hintikka, E.; Pihlajaniemi, T.
J. Biol. Chem. 269, 13929-13935, 1994
A:Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial collagen chain.
A:Reference number: A54072; MUID:94245707
A:Accession: A54072
A:Molecule type: DNA; mRNA
A:Residues: 1293-1403, 'R', 1405-1774 <REH3>
A:Cross-references: GB:U03714; NID:g487733; PIDN:AAA20657.1; PID:g487734
R;O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Cell 88, 277-285, 1997
A:Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
A:Reference number: A58816; MUID:97160848
A:Accession: A58816
A:Molecule type: protein
A:Residues: 1591-1610 <ORE>
A:Experimental source: hemangioendothelium cells
A:Note: Inhibits endothelial cell proliferation
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit and subsequently O-glycosylated.
C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in p
C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of
ay be useful in treating solid tumors.
C:Genetics:

A:Gene: MGI:Coll18a1
A:Cross-references: MGI:71175
A:Map position: 10:41.0
A:Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 15
A:Note: the list of introns is incomplete
C:Superfamily: unassigned collagens
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog
F;1-1774/Product: collagen alpha 1(XVII) chain precursor, long splice form #status p
F;1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form
F;1-24/Domain: signal sequence #status predicted <Sig>
F;361-486/Region: frizzled similarity
F;786-812/Domain: collagenous #status predicted <CO01>
F;823-896/Domain: collagenous #status predicted <CO02>
F;921-1042/Domain: collagenous #status predicted <CO03>
F;1066-1148/Domain: collagenous #status predicted <CO04>
F;1163-1204/Domain: collagenous #status predicted <CO05>
F;1218-1290/Domain: collagenous #status predicted <CO06>
F;1301-1333/Domain: collagenous #status predicted <CO07>
F;1346-1369/Domain: collagenous #status predicted <CO08>
F;1351-1353/Region: cell attachment (R-G-D) motif
F;1377-1428/Domain: collagenous #status predicted <CO09>
F;1442-1459/Domain: collagenous #status predicted <CO10>
F;1591-1774/Product: endostatin #status predicted <EST>
F;1598-1774/Region: multiplexin collagen carboxyl-terminal similarity
F;354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 45; DB 2; Length 1774;
Best Local Similarity 100.0%; Pred. No. 0.51;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||||

Db 1766 NSFMTSFSK 1774

RESULT 3
T26721

hypothetical protein Y39A1A.20 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26721
R;Wall, M.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z20257
A:Accession: T26721
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-128 <WIL>
A:Cross-references: EMBL:AL031633; PIDN:CAA21010.1; GSPDB:GN00021; CESP:Y39A1A.20
A:Experimental source: clone Y39A1A
C:Genetics:
A:Gene: CESP:Y39A1A.20
A:Map position: 3
A:Introns: 33/2; 47/1; 77/3; 107/1

Query Match 75.6%; Score 34; DB 2; Length 128;
Best Local Similarity 66.7%; Pred. No. 7.1;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||:||||
Db 7 NSFLVSFSR 15

RESULT 4
QQBEIS

UL89 protein - human cytomegalovirus (strain AD169)
C:Species: human cytomegalovirus, human herpesvirus 5
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 17-Feb-1995
C:Accession: S09853
R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.;
M.; Barrell, B.G.
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990
A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus
A:Reference number: S09749; MUID:90269039
A:Accession: S09853
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-724 <CHE>
A:Cross-references: EMBL:X17403
A:Note: possible protein-coding frames are given
A:Note: the DNA sequence was submitted to EMBL, December 1989, in computer-readable form
C:Superfamily: herpesvirus 38K protein

Query Match 75.6%; Score 34; DB 1; Length 724;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||:||||
Db 128 NFMWTFKK 136

RESULT 5
F95011

conserved hypothetical protein SP0100 [imported] - Streptococcus pneumoniae (strain TIGR
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 22-Oct-2001
C:Accession: F95011
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95011
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-108 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74287.1; PID:g14971566; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0100
C:Superfamily: conserved hypothetical protein MTH281

Query Match 73.3%; Score 33; DB 2; Length 108;
Best Local Similarity 55.6%; Pred. No. 9.7;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||:||||
Db 53 NSFLTYSR 61

RESULT 6
A97883

conserved hypothetical protein spr0089 [imported] - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: A97883
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: A97883
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-108 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK98893.1; PID:g15457624; GSPDB:GN00174
C:Genetics:
A:Gene: spr0089
C:Superfamily: conserved hypothetical protein MTH281

Query Match 73.3%; Score 33; DB 2; Length 108;
Best Local Similarity 55.6%; Pred. No. 9.7;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||:||||
Db 53 NSFLTYSR 61

RESULT 7
T20115

hypothetical protein C50F4.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20115
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19225
A:Accession: T20115
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-300 <WIL>
A:Cross-references: EMBL:Z70750; PIDN:CAA94743.1; GSPDB:GN00023; CESP:C50F4.8
A:Experimental source: clone C50F4
C:Genetics:
A:Gene: CESP:C50F4.8
A:Map position: 5
A:Introns: 32/2; 90/2

Query Match 73.3%; Score 33; DB 2; Length 300;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFS 8
|||:||||
Db 73 NDFLTSPS 80

RESULT 8

JQ0699

hypothetical 19K protein (secD region) - Escherichia coli

N:Alternate names: hypothetical protein 7

C:Species: Escherichia coli

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 04-Mar-2000

C:Accession: JQ0699

R:Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.

EMBO J. 9, 3209-3216, 1990

A:Title: The secD locus of E.coli codes for two membrane proteins required for protein e

A:Reference number: JQ0693; MUID:91006014

A:Accession: JQ0699

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-171 <GAR>

C:Superfamily: Escherichia coli hypothetical 19K protein (secD region)

Query Match 71.1%; Score 32; DB 2; Length 171;

Best Local Similarity 75.0%; Pred. No. 25;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFS 8

||:||||

Db 136 NGFITSFS 143

RESULT 9

T31621

hypothetical protein Y50E8A.q - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Aug-2000

C:Accession: T31621

R:Steward, C.

submitted to the EMBL Data Library, September 1999

A:Reference number: 221047

A:Accession: T31621

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-183 <WIL>

A:Cross-references: EMBL:AL117200; NID:el549770; PIDN:CAB55060.1; CESP:Y50E8A.q

A:Experimental source: clone Y50E8A

C:Genetics:

A:Gene: CESP:Y50E8A.q

A:Introns: 42/2; 101/3; 127/3

C:Superfamily: Caenorhabditis elegans hypothetical protein Y50E8A.q

Query Match

Best Local Similarity 71.1%; Score 32; DB 2; Length 183;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9

||:||||:

Db 29 NSIVTSFTR 37

RESULT 10

H97015

glutamine amidotransferase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: H97015

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: H97015

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-203 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK78915.1; PID:gl5023841; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC0939

C:Superfamily: amidotransferase hisH; trpG homology

Query Match

Best Local Similarity 71.1%; Score 32; DB 2; Length 203;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9

||:||||

Db 28 NSFITSDSK 36

RESULT 11

A61073

heat shock protein 90 homolog - yeast (Candida albicans) (fragment)

N:Alternate names: 47K antigen

C:Species: Candida albicans

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Nov-1996

C:Accession: A61073

R:Matthews, R.; Burnie, J.

FEMS Microbiol. Lett. 60, 25-30, 1989

A:Title: Cloning of a DNA sequence encoding a major fragment of the 47 kilodalton str

A:Reference number: A61073

A:Accession: A61073

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-395 <MAT>

C:Comment: Candida albicans is an important human fungal pathogen. This protein is a

C:Superfamily: heat shock protein 90

C:Keywords: heat shock; stress-induced protein

Query Match

Best Local Similarity 71.1%; Score 32; DB 2; Length 395;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9

||:||||

Db 100 NQFYTAFSK 108

RESULT 12

A53234

globulin-1S, GLB1S - maize

C:Species: Zea mays (maize)

C:Date: 02-May-1994 #sequence_revision 18-Nov-1994 #text_change 11-Jan-2000

C:Accession: A53234; A43642

R:Belanger, F.C.; Kriz, A.L.

Genetics 129, 863-872, 1991

A:Title: Molecular basis for allelic polymorphism of the maize Globulin-1 gene.

A:Reference number: A53234; MUID:92090707

A:Accession: A53234

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-573 <BEL>

A:Experimental source: inbred line Va 26

A:Note: sequence extracted from NCBI backbone (NCBIN:71280, NCBIIP:71284)

R:Kriz, A.L.

Biochem. Genet. 27, 239-251, 1989

A:Title: Characterization of embryo globulins encoded by the maize Glb genes.

A:Reference number: A43642; MUID:89374022

A:Accession: A43642

A:Status: preliminary

A:Molecule type: protein

A:Residues: 87-100 <KRI>

C:Superfamily: glycinin

Query Match

Best Local Similarity 71.1%; Score 32; DB 2; Length 573;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFMTSFSK 9
|||||
Db 240 SFLSSFSK 247

RESULT 13
B53234
vicilin-like storage protein Gbl1-L, embryo - maize
N;Alternate names: globulin-1L
C;Species: zea mays (maize)
C;Date: 02-May-1994 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
C;Accession: B53234; S21824
R;Belanger, F.C.; Kriz, A.L.
Genetics 129, 863-872, 1991
A;Title: Molecular basis for allelic polymorphism of the maize Globulin-1 gene.
A;Reference number: A53234; MUID:92090707
A;Accession: B53234
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-582 <BEL>
A;Cross-references: EMBL:X59083; NID:g22283; PIDN:CAA41809.1; PID:g22284
A;Experimental source: inbred line W64A6
A;Note: sequence extracted from NCBI backbone (NCBIP:71285)
C;Genetics:
A;Gene: Gbl1-L
A;Introns: 167/1; 225/3; 252/3; 349/3
C;Superfamily: glycinin

Query Match 71.1%; Score 32; DB 2; Length 582;
Best Local Similarity 75.0%; Pred. No. 84;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFMTSFSK 9
|||||
Db 237 SFLSSFSK 244

RESULT 14
H96692
probable receptor serine/threonine kinase PR5K T4024.8 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H96692
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: H96692
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-655 <STO>
A;Cross-references: GB:AE005173; NID:gl1128390; PIDN:AAG31195.1; GSPDB:GN00141
C;Genetics:
A;Gene: T4024.8
A;Map position: 1

Query Match 71.1%; Score 32; DB 2; Length 655;
Best Local Similarity 66.7%; Pred. No. 94;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||
Db 101 NSFMTSFNE 109

RESULT 15
T28636
zinc-metalloproteinase-like protein - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T28636
R;Florent, I.C.P.; Derhy, Z.; Allary, M.; Monsigny, M.; Mayer, R.; Schrevel, J.
Mol. Biochem. Parasitol. 97, 149-160, 1998
A;Title: A Plasmodium falciparum aminopeptidase gene belonging to the M1 family of z
A;Reference number: Z20491; MUID:99094503
A;Accession: T28636
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1056 <FLO>
A;Cross-references: EMBL:Y09081; NID:el361211; PID:el361212; PIDN:CAA70301.1

Query Match 71.1%; Score 32; DB 2; Length 1056;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||
Db 243 NEFLTIFSK 251

Search completed: May 31, 2002, 10:28:25
Job time: 167 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:32:17 ; Search time 15.74 Seconds
(without alignments)
17.247 Million cell updates/sec

Title: US-09-589-777C-25
Perfect score: 45
Sequence: 1 NSFMTSFSK 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 124036 seqs, 30162252 residues

Total number of hits satisfying chosen parameters: 124036

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	73.3	556	5	US-09-573-655B-477 Sequence 477, App
2	33	73.3	556	5	US-09-573-655B-558 Sequence 558, App
3	33	73.3	556	5	US-09-573-655B-1894 Sequence 1894, Ap
4	32	71.1	1335	5	US-09-540-209B-7531 Sequence 7531, Ap
5	31	68.9	314	5	US-09-602-777A-74 Sequence 74, Appl
6	31	68.9	1119	6	US-10-121-049-352 Sequence 352, App
7	31	68.9	1119	6	US-10-121-050-352 Sequence 352, App
8	31	68.9	1119	6	US-10-121-053-352 Sequence 352, App
9	31	68.9	1119	6	US-10-121-043-352 Sequence 352, App
10	31	68.9	1119	6	US-10-121-044-352 Sequence 352, App
11	31	68.9	1119	6	US-10-121-047-352 Sequence 352, App
12	31	68.9	1119	6	US-10-121-054-352 Sequence 352, App
13	31	68.9	1119	6	US-10-121-056-352 Sequence 352, App
14	31	68.9	1119	6	US-10-121-057-352 Sequence 352, App
15	31	68.9	1119	6	US-10-121-058-352 Sequence 352, App
16	31	68.9	1119	6	US-10-121-060-352 Sequence 352, App
17	31	68.9	1119	6	US-10-121-063-352 Sequence 352, App
18	31	68.9	1119	6	US-10-123-108-352 Sequence 352, App
19	31	68.9	1119	6	US-10-123-154-352 Sequence 352, App
20	31	68.9	1119	6	US-10-123-156-352 Sequence 352, App
21	31	68.9	1119	6	US-10-123-157-352 Sequence 352, App
22	31	68.9	1119	6	US-10-123-212-352 Sequence 352, App
23	31	68.9	1119	6	US-10-123-213-352 Sequence 352, App
24	31	68.9	1119	6	US-10-123-109-352 Sequence 352, App
25	31	68.9	1119	6	US-10-121-041-352 Sequence 352, App
26	31	68.9	1119	6	US-10-121-045-352 Sequence 352, App

27	31	68.9	1119	6	US-10-121-046-352	Sequence 352, App
28	31	68.9	1119	6	US-10-121-051-352	Sequence 352, App
29	31	68.9	1119	6	US-10-121-040-352	Sequence 352, App
30	31	68.9	1119	6	US-10-121-048-352	Sequence 352, App
31	31	68.9	1119	6	US-10-121-052-352	Sequence 352, App
32	31	68.9	1119	6	US-10-121-061-352	Sequence 352, App
33	31	68.9	1119	6	US-10-121-042-352	Sequence 352, App
34	31	68.9	1119	6	US-10-121-055-352	Sequence 352, App
35	31	68.9	1119	6	US-10-121-059-352	Sequence 352, App
36	31	68.9	1119	6	US-10-124-822-352	Sequence 352, App
37	31	68.9	1119	6	US-10-123-903-352	Sequence 352, App
38	31	68.9	1119	6	US-10-124-817-352	Sequence 352, App
39	31	68.9	1119	6	US-10-124-819-352	Sequence 352, App
40	31	68.9	1119	6	US-10-124-823-352	Sequence 352, App
41	31	68.9	1119	6	US-10-125-704-352	Sequence 352, App
42	31	68.9	1119	6	US-10-123-215-352	Sequence 352, App
43	31	68.9	1119	6	US-10-123-235-352	Sequence 352, App
44	31	68.9	1119	6	US-10-123-236-352	Sequence 352, App
45	31	68.9	1119	6	US-10-123-261-352	Sequence 352, App

ALIGNMENTS

RESULT 1
US-09-573-655B-477
; Sequence 477, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptic
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573.655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 477
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-477

Query Match 73.3%; Score 33; DB 5; Length 556;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NSFMTSFSK 9
|||:|:|
Db 149 NSFIASYSK 157

RESULT 2
US-09-573-655B-558
; Sequence 558, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptic
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573.655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 558
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-558

Query Match 73.3%; Score 33; DB 5; Length 556;
Best Local Similarity 66.7%; Pred. No. 30;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||:|:|
Db 149 NSFIASYSK 157

RESULT 3

US-09-573-655B-1894
; Sequence 1894, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573.655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1894
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-1894

Query Match 73.3%; Score 33; DB 5; Length 556;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||:|:|
Db 149 NSFIASYSK 157

RESULT 4

US-09-540-209B-7531
; Sequence 7531, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540.209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 7531
; LENGTH: 1335
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-7531

Query Match 71.1%; Score 32; DB 5; Length 1335;
Best Local Similarity 66.7%; Pred. No. 1.le+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
:|:|:|:|
Db 386 HSFMSSEK 394

RESULT 5

US-09-602-777A-74
; Sequence 74, Application US/09602777A
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN HOMEOSTASIS AND ADAPTATION

; FILE REFERENCE: BGI-128CP
; CURRENT APPLICATION NUMBER: US/09/602,777A
; CURRENT FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931636.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932127.2
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932129.9
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: DE 19932226.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932920.6
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932922.2
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932924.9
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932928.1
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932930.3
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932933.8
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932935.4
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932973.7
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933002.6
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933003.4
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933005.0
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933006.9
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19941378.9
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941379.7
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941390.8
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941391.6
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19942088.2
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 442
; SEQ ID NO 74
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-602-777A-74

Query Match 68.9%; Score 31; DB 5; Length 314;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|:|:|:|
Db 79 NSFMTALHK 87

RESULT 6

US-10-121-049-352
; Sequence 352, Application US/10121049

```

; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 352
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-049-352
```

Query Match 68.9%; Score 31; DB 6; Length 1119;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
1111 :| |
Db 1035 NSFMTGTFGK 1043

```

RESULT 7
US-10-121-050-352
; Sequence 352, Application US/10121050
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C20
; CURRENT APPLICATION NUMBER: US/10/121,050
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 352
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Homo Sapien
```

US-10-121-050-352

Query Match 68.9%; Score 31; DB 6; Length 1119;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
1111 :| |
Db 1035 NSFMTGTFGK 1043

```

RESULT 8
US-10-121-053-352
; Sequence 352, Application US/10121053
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C23
; CURRENT APPLICATION NUMBER: US/10/121,053
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 352
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-053-352
```

Query Match 68.9%; Score 31; DB 6; Length 1119;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
1111 :| |
Db 1035 NSFMTGTFGK 1043

```

RESULT 9
US-10-121-043-352
; Sequence 352, Application US/10121043
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
```



```
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C15
; CURRENT APPLICATION NUMBER: US/10/121,043
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 352
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-043-352
```

```
Query Match 68.9%; Score 31; DB 6; Length 1119;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 NSFMTSFSK 9
    ||||:|
Db 1035 NSFMTGTFGK 1043
```

```
RESULT 10
US-10-121-044-352
; Sequence 352, Application US/10121044
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C5
; CURRENT APPLICATION NUMBER: US/10/121,044
; CURRENT FILING DATE: 2002-04-11
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 352
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-044-352
```

```
Query Match 68.9%; Score 31; DB 6; Length 1119;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 NSFMTSFSK 9
    ||||:|
Db 1035 NSFMTGTFGK 1043
```

```
RESULT 11
US-10-121-047-352
```

```
; Sequence 352, Application US/10121047
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C4
; CURRENT APPLICATION NUMBER: US/10/121,047
; CURRENT FILING DATE: 2002-04-11
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 352
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-047-352
```

```
Query Match 68.9%; Score 31; DB 6; Length 1119;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 NSFMTSFSK 9
    ||||:|
Db 1035 NSFMTGTFGK 1043
```

```
RESULT 12
US-10-121-054-352
; Sequence 352, Application US/10121054
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C14
; CURRENT APPLICATION NUMBER: US/10/121,054
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 352
; LENGTH: 1119
; TYPE: PRT
```

; ORGANISM: Homo Sapien
US-10-121-054-352

Query Match 68.9%; Score 31; DB 6; Length 1119;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
| | | | : | |
Db 1035 NSFMTGTFGK 1043

RESULT 13
US-10-121-056-352
; Sequence 352, Application US/10121056
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC25
; CURRENT APPLICATION NUMBER: US/10/121,056
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 352
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-056-352

Query Match 68.9%; Score 31; DB 6; Length 1119;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
| | | | : | |
Db 1035 NSFMTGTFGK 1043

RESULT 14
US-10-121-057-352
; Sequence 352, Application US/10121057
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC13
; CURRENT APPLICATION NUMBER: US/10/121,057
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 352
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-057-352

Query Match 68.9%; Score 31; DB 6; Length 1119;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
| | | | : | |
Db 1035 NSFMTGTFGK 1043

RESULT 15
US-10-121-058-352
; Sequence 352, Application US/10121058
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC12
; CURRENT APPLICATION NUMBER: US/10/121,058
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 352
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-058-352

Query Match 68.9%; Score 31; DB 6; Length 1119;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
| | | | : | |
Db 1035 NSFMTGTFGK 1043

Search completed: May 31, 2002, 10:32:18
Job time: 350 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:31:55 ; Search time 205.32 Seconds
(without alignments)
15.429 Million cell updates/sec

Title: US-09-589-777C-25
Perfect score: 45
Sequence: 1 NSFMTSRSK 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*

2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*

3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*

4: /cgn2_6/ptodata/2/paa/US080_COMB.pep:*

5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*

6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*

7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*

8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*

9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*

10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*

11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*

12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*

13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*

14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*

15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*

16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*

17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*

18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*

19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*

20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*

21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*

22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*

23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*

24: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*

25: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*

26: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	9	19	US-09-589-777C-25
2	45	100.0	50	23	US-09-958-489-4
3	45	100.0	85	19	US-09-589-774-5
4	45	100.0	184	1	PCT-US00-12063-4
5	45	100.0	184	1	PCT-US98-24950-10
6	45	100.0	184	13	US-08-975-424-10
7	45	100.0	184	17	US-09-353-333-4
					Sequence 25, Appli
					Sequence 4, Appli
					Sequence 5, Appli
					Sequence 4, Appli
					Sequence 10, Appl
					Sequence 4, Appli

8	45	100.0	184	17	US-09-383-315-18	Sequence 18, Appl
9	45	100.0	184	19	US-09-589-777A-2	Sequence 2, Appli
10	45	100.0	184	19	US-09-589-777C-2	Sequence 2, Appli
11	45	100.0	191	19	US-09-561-005-13	Sequence 13, Appl
12	45	100.0	191	19	US-09-561-499-13	Sequence 13, Appl
13	45	100.0	191	19	US-09-561-526-13	Sequence 13, Appl
14	45	100.0	191	19	US-09-562-245-13	Sequence 13, Appl
15	45	100.0	191	23	US-09-998-831-13	Sequence 13, Appl
16	45	100.0	207	19	US-09-586-625-71	Sequence 71, Appl
17	45	100.0	207	24	US-10-080-797-3	Sequence 3, Appli
18	45	100.0	218	1	PCT-US98-24950-14	Sequence 14, Appl
19	45	100.0	218	13	US-08-975-424-14	Sequence 14, Appl
20	45	100.0	580	1	PCT-US98-24950-16	Sequence 16, Appl
21	45	100.0	580	13	US-08-975-424-16	Sequence 16, Appl
22	42	93.3	185	24	US-10-036-869-36	Sequence 36, Appl
23	38	84.4	184	17	US-09-383-315-35	Sequence 35, Appl
24	34	75.6	83	21	US-09-758-465-1129	Sequence 1129, Ap
25	34	75.6	337	16	US-09-248-796-15700	Sequence 15700, A
26	34	75.6	337	26	US-60-096-409-15700	Sequence 15700, A
27	33	73.3	75	15	US-09-107-433-2800	Sequence 2800, Ap
28	33	73.3	108	1	PCT-US97-14436-553	Sequence 553, App
29	33	73.3	108	13	US-08-911-503-553	Sequence 553, App
30	33	73.3	108	13	US-08-911-503A-553	Sequence 553, App
31	33	73.3	108	19	US-09-583-110-4993	Sequence 4993, Ap
32	33	73.3	117	15	US-09-107-433-3291	Sequence 3291, Ap
33	33	73.3	124	19	US-09-583-110-4960	Sequence 4960, Ap
34	33	73.3	143	15	US-09-107-433-2801	Sequence 2801, Ap
35	33	73.3	556	19	US-09-573-655A-477	Sequence 477, App
36	33	73.3	556	19	US-09-573-655A-558	Sequence 558, App
37	33	73.3	556	19	US-09-573-655A-1894	Sequence 1894, Ap
38	33	73.3	792	19	US-09-570-581A-1212	Sequence 1212, Ap
39	33	73.3	3287	12	US-08-851-486-7	Sequence 7, Appli
40	33	73.3	3287	17	US-09-304-533-7	Sequence 7, Appli
41	32	71.1	60	1	PCT-US01-01354-10056	Sequence 10056, A
42	32	71.1	60	21	US-09-764-905-10056	Sequence 10056, A
43	32	71.1	60	24	US-10-092-399-10056	Sequence 10056, A
44	32	71.1	246	1	PCT-US98-06371-108	Sequence 108, App
45	32	71.1	246	12	US-08-833-457-108	Sequence 108, App

ALIGNMENTS

RESULT 1
US-09-589-777C-25
; Sequence 25, Application US/09589777C
; GENERAL INFORMATION:
; APPLICANT: Sukhame, Vikas P.
; TITLE OF INVENTION: Anti-Angiogenic Peptides and Methods of
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: 1440.1023-011
; CURRENT APPLICATION NUMBER: US/09/589,777C
; CURRENT FILING DATE: 2000-06-08
; PRIOR APPLICATION NUMBER: PCT/US98/26057
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 60/108,536
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: US 60/082,663
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: US 60/067,888
; PRIOR FILING DATE: 1997-12-08
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-589-777C-25

Query Match 100.0%; Score 45; DB 19; Length 9;
Best Local Similarity 100.0%; Pred. No. 3.2e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 NSFMTSFSK 9
    |||||
Db 1 NSFMTSFSK 9

RESULT 2
US-09-958-489-4
; Sequence 4, Application US/09958489
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITA' DEGLI STUDI DI MILANO
; APPLICANT: UNIVERSITA' DEGLI STUDI DI FIRENZE
; TITLE OF INVENTION: POLYPEPTIDES WITH ANTIANGIOGENIC ACTIVITY
; FILE REFERENCE: UNIVERSITA'
; CURRENT APPLICATION NUMBER: US/09/958,489
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (39)
; OTHER INFORMATION: T-BUTYL
; OTHER INFORMATION: Description of Artificial Sequence: polypeptides
; OTHER INFORMATION: homologous to endostatin
; NAME/KEY: DISULFID
; LOCATION: (1)..(31)
; OTHER INFORMATION: intramolecular disulfidic bond
US-09-958-489-4

Query Match 100.0%; Score 45; DB 23; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.33; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 NSFMTSFSK 9
    |||||
Db 42 NSFMTSFSK 50

RESULT 3
US-09-589-774-5
; Sequence 5, Application US/09589774
; GENERAL INFORMATION:
; APPLICANT: Sukhatme, Vikas P.
; TITLE OF INVENTION: Restin and Methods of Use Thereof
; FILE REFERENCE: 1440.1014011
; CURRENT APPLICATION NUMBER: US/09/589,774
; CURRENT FILING DATE: 2000-06-08
; PRIOR APPLICATION NUMBER: PCT/US98/26058
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: 60/108,536
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 60/082,663
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/067,888
; PRIOR FILING DATE: 1997-12-08
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-589-774-5

Query Match 100.0%; Score 45; DB 19; Length 85;
Best Local Similarity 100.0%; Pred. No. 0.56; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 NSFMTSFSK 9
    |||||
Db 77 NSFMTSFSK 85

RESULT 4
PCT-US00-12063-4
; Sequence 4, Application PC/TUS00012063
; GENERAL INFORMATION:
; APPLICANT: The Burnham Institute
; TITLE OF INVENTION: Antiangiogenic Endostatin Peptides, Endostatin Variants
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: FP-LJ 4148
; CURRENT APPLICATION NUMBER: PCT/US00/12063
; CURRENT FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: US 09/353,333
; PRIOR FILING DATE: 1999-07-14
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
PCT-US00-12063-4

Query Match 100.0%; Score 45; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.2; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 NSFMTSFSK 9
    |||||
Db 176 NSFMTSFSK 184

RESULT 5
PCT-US98-24950-10
; Sequence 10, Application PC/TUS9824950
; GENERAL INFORMATION:
; APPLICANT: Genetix Pharmaceuticals, Inc.
; TITLE OF INVENTION: ANTI-ANGIOGENIC GENE THERAPY VECTORS AND
; TITLE OF INVENTION: THEIR USE IN TREATING ANGIOGENESIS-RELATED DISEASES
; FILE REFERENCE: 50033/002W01
; CURRENT APPLICATION NUMBER: PCT/US98/24950
; CURRENT FILING DATE: 1998-12-29
; EARLIER APPLICATION NUMBER: 08/975,424
; EARLIER FILING DATE: 1997-11-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
PCT-US98-24950-10

Query Match 100.0%; Score 45; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.2; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 NSFMTSFSK 9
    |||||
Db 176 NSFMTSFSK 184

RESULT 6
US-08-975-424-10
; Sequence 10, Application US/08975424
; GENERAL INFORMATION:
; APPLICANT: LeBoulch, Philipp
; TITLE OF INVENTION: ANTI-ANGIOGENIC GENE THERAPY VECTORS AND
; TITLE OF INVENTION: THEIR USE IN TREATING ANGIOGENESIS-RELATED DISEASES
; FILE REFERENCE: 50033/002001
```

; CURRENT APPLICATION NUMBER: US/08/975,424
; CURRENT FILING DATE: 1997-11-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-975-424-10

Query Match 100.0%; Score 45; DB 13; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 176 NSFMTSFSK 184

RESULT 7
US-09-353-333-4
; Sequence 4, Application US/09353333A
; GENERAL INFORMATION:
; APPLICANT: Vuori, Kristiina
; TITLE OF INVENTION: Antiangiogenic Endostatin Peptides, Endostatin Variants
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: P-LJ 3557
; CURRENT APPLICATION NUMBER: US/09/353,333A
; CURRENT FILING DATE: 1999-07-14
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-353-333-4

Query Match 100.0%; Score 45; DB 17; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 176 NSFMTSFSK 184

RESULT 8
US-09-383-315-18
; Sequence 18, Application US/093833315A
; GENERAL INFORMATION:
; APPLICANT: Lo, Kin-Ming
; APPLICANT: Li, Yue
; APPLICANT: Gillies, Stephen D
; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as
; TITLE OF INVENTION: Immunofusins
; FILE REFERENCE: LEX-006
; CURRENT APPLICATION NUMBER: US/09/383,315A
; CURRENT FILING DATE: 1999-08-25
; EARLIER APPLICATION NUMBER: US 60/097,883
; EARLIER FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-383-315-18

Query Match 100.0%; Score 45; DB 17; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NSFMTSFSK 9
Db 176 NSFMTSFSK 184

RESULT 9
US-09-589-777A-2
; Sequence 2, Application US/09589777A
; GENERAL INFORMATION:
; APPLICANT: Sukhatme, Vikas P.
; TITLE OF INVENTION: Anti-Angiogenic Peptides and Methods of
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: 1440.1023-011
; CURRENT APPLICATION NUMBER: US/09/589,777A
; CURRENT FILING DATE: 2000-06-08
; PRIOR APPLICATION NUMBER: PCT/US98/26057
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: US 60/108,536
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: US 60/082,663
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: US 60/067,888
; PRIOR FILING DATE: 1997-12-07
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-589-777A-2

Query Match 100.0%; Score 45; DB 19; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 176 NSFMTSFSK 184

RESULT 10
US-09-589-777C-2
; Sequence 2, Application US/09589777C
; GENERAL INFORMATION:
; APPLICANT: Sukhatme, Vikas P.
; TITLE OF INVENTION: Anti-Angiogenic Peptides and Methods of
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: 1440.1023-011
; CURRENT APPLICATION NUMBER: US/09/589,777C
; CURRENT FILING DATE: 2000-06-08
; PRIOR APPLICATION NUMBER: PCT/US98/26057
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 60/108,536
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: US 60/082,663
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: US 60/067,888
; PRIOR FILING DATE: 1997-12-08
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-589-777C-2

Query Match 100.0%; Score 45; DB 19; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||

Db 176 NSFMTSFSK 184

RESULT 11

US-09-561-005-13

; Sequence 13, Application US/09561005

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Rolf A. Brekken

; TITLE OF INVENTION: ANTIBODY CONJUGATE METHODS FOR SELECTIVELY INHIBITING VEGF

; FILE REFERENCE: 4001.002585

; CURRENT APPLICATION NUMBER: US/09/561,005

; CURRENT FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/131,432

; PRIOR FILING DATE: 1999-04-28

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 191

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

US-09-561-005-13

Query Match 100.0%; Score 45; DB 19; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9

|||||

Db 183 NSFMTSFSK 191

RESULT 12

US-09-561-499-13

; Sequence 13, Application US/09561499

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Rolf A. Brekken

; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF

; FILE REFERENCE: 4001.002582

; CURRENT APPLICATION NUMBER: US/09/561,499

; CURRENT FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/131,432

; PRIOR FILING DATE: 1999-04-28

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 191

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

US-09-561-499-13

Query Match

100.0%; Score 45; DB 19; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9

|||||

Db 183 NSFMTSFSK 191

RESULT 13

US-09-561-526-13

; Sequence 13, Application US/09561526

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Rolf A. Brekken

; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF

; FILE REFERENCE: 4001.002586

; CURRENT APPLICATION NUMBER: US/09/561,526

; CURRENT FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/131,432

; PRIOR FILING DATE: 1999-04-28

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 191

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

US-09-561-526-13

Query Match

100.0%; Score 45; DB 19; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9

|||||

Db 183 NSFMTSFSK 191

RESULT 14

US-09-562-245-13

; Sequence 13, Application US/09562245

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Rolf A. Brekken

; TITLE OF INVENTION: ANTIBODY KITS FOR SELECTIVELY INHIBITING VEGF

; FILE REFERENCE: 4001.002583

; CURRENT APPLICATION NUMBER: US/09/562,245

; CURRENT FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/131,432

; PRIOR FILING DATE: 1999-04-28

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 191

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

US-09-562-245-13

Query Match

100.0%; Score 45; DB 19; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9

|||||

Db 183 NSFMTSFSK 191

RESULT 15

US-09-998-831-13

; Sequence 13, Application US/09998831

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Rolf A. Brekken

; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY

; TITLE OF INVENTION: INHIBITING VEGF

; FILE REFERENCE: 4001.002584

; CURRENT APPLICATION NUMBER: US/09/998,831

; CURRENT FILING DATE: 2001-11-30

; PRIOR APPLICATION NUMBER: 09/561,108


```
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-998-831-13

Query Match      100.0%; Score 45; DB 23; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NSPMTSFSK 9
        |||||
Db      183 NSPMTSFSK 191
```

Search completed: May 31, 2002, 10:31:56
Job time: 363 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:27:50 ; Search time 21.9 Seconds
(without alignments)
10.038 Million cell updates/sec

Title: US-09-589-777C-25
Perfect score: 45
Sequence: 1 NSFMTSFSK 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	191	4	US-09-561-500-13
2	45	100.0	191	4	US-09-561-108-13
3	45	100.0	195	1	US-08-159-784-2
4	42	93.3	185	3	US-08-985-526-36
5	33	73.3	3287	2	US-08-477-451-7
6	32	71.1	395	1	US-08-357-264-1
7	32	71.1	395	1	US-08-672-514-1
8	30	66.7	274	3	US-09-141-821-3
9	30	66.7	285	3	US-09-141-821-1
10	30	66.7	285	3	US-09-141-821-2
11	30	66.7	285	3	US-09-141-821-4
12	30	66.7	285	3	US-09-141-821-5
13	29	64.4	139	1	US-08-330-978-2
14	29	64.4	139	1	US-08-474-042-2
15	29	64.4	139	1	US-08-484-558-2
16	29	64.4	139	1	US-08-774-592-2
17	29	64.4	286	1	US-08-202-186-25
18	29	64.4	286	2	US-08-418-071-18
19	29	64.4	437	1	US-08-487-037-1
20	29	64.4	437	1	US-08-487-037-3
21	29	64.4	488	1	US-08-487-037-1
22	28	62.2	191	1	US-08-159-784-3
23	28	62.2	262	2	US-08-038-761A-1
24	28	62.2	321	4	US-09-081-686-4
25	28	62.2	446	4	US-09-081-686-2
26	28	62.2	582	2	US-08-989-386-1
27	28	62.2	1009	2	US-08-680-326-31

28	62.2	2183	1	US-08-348-891A-7	Sequence 7, Appli
29	62.2	2183	2	US-08-905-817-7	Sequence 7, Appli
30	60.0	129	4	US-09-063-743-3	Sequence 3, Appli
31	60.0	178	4	US-09-315-689-5	Sequence 5, Appli
32	60.0	182	4	US-09-561-500-14	Sequence 14, Appl
33	60.0	182	4	US-09-561-108-14	Sequence 14, Appl
34	60.0	182	4	US-09-315-689-3	Sequence 3, Appli
35	60.0	183	4	US-09-206-059-2	Sequence 2, Appli
36	60.0	331	2	US-08-878-989-21	Sequence 21, Appl
37	60.0	331	3	US-09-101-146-64	Sequence 64, Appl
38	60.0	331	4	US-09-272-796-21	Sequence 21, Appl
39	60.0	411	1	US-08-381-433A-4	Sequence 4, Appli
40	60.0	411	1	US-08-381-433A-8	Sequence 8, Appli
41	60.0	411	4	US-08-981-189B-12	Sequence 12, Appl
42	60.0	415	1	US-08-110-286A-2	Sequence 2, Appli
43	60.0	415	1	US-08-110-286A-6	Sequence 6, Appli
44	60.0	415	4	US-08-981-189B-10	Sequence 10, Appl
45	60.0	431	1	US-08-381-433A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-561-500-13
; Sequence 13, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-500-13

Query Match 100.0%; Score 45; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
| | | | | | | | | |
Db. 183 NSFMTSFSK 191

RESULT 2
US-09-561-108-13
; Sequence 13, Application US/095611108
; Patent No. 6342221
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VE
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-108-13

Query Match      100.0%; Score 45; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 183 NSFMTSFSK 191

RESULT 3
US-08-159-784-2
; Sequence 2, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
US-08-159-784-2

Query Match      100.0%; Score 45; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 187 NSFMTSFSK 195

RESULT 4
US-08-985-526-36
; Sequence 36, Application US/08985526
; Patent No. 6080728
```

```

; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 185 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-985-526-36

Query Match      93.3%; Score 42; DB 3; Length 185;
Best Local Similarity 88.9%; Pred. No. 0.17;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 177 NSFMTSFSK 185

RESULT 5
US-08-477-451-7
; Sequence 7, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
```

REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3287 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-451-7

Query Match 73.3%; Score 33; DB 2; Length 3287;
Best Local Similarity 77.8%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
| | | | |
Db 1117 NFFSFSK 1125

RESULT 6
US-08-357-264-1
; Sequence 1, Application US/08357264
; Patent No. 5541077
; GENERAL INFORMATION:
; APPLICANT: BURNIE Mr., James P.
; APPLICANT: MATTHEWS Ms., Ruth C.
; TITLE OF INVENTION: FUNGAL STRESS PROTEINS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cushman, Darby & Cushman
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/357,264
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/152669
; FILING DATE: 16-NOV-1993
; APPLICATION NUMBER: US 663897
; FILING DATE: 14-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, Paul N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: PNK/5544/202253/DJP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-357-264-1

Query Match 71.1%; Score 32; DB 1; Length 395;

Best Local Similarity 66.7%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 NSFMTSFSK 9
| | | | |
Db 100 NQFYTAFSK 108
RESULT 7
US-08-672-514-1
; Sequence 1, Application US/08672514
; Patent No. 5686248
; GENERAL INFORMATION:
; APPLICANT: BURNIE Mr., James P.
; APPLICANT: MATTHEWS Ms., Ruth C.
; TITLE OF INVENTION: FUNGAL STRESS PROTEINS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cushman, Darby & Cushman
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/672,514
; FILING DATE: 28-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/152,669
; FILING DATE: 16-NOV-1993
; APPLICATION NUMBER: US 08/152669
; FILING DATE: 16-NOV-1993
; APPLICATION NUMBER: US 663897
; FILING DATE: 14-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, Paul N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: PNK/5544/202253/DJP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-672-514-1

Query Match 71.1%; Score 32; DB 1; Length 395;
Best Local Similarity 66.7%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 NSFMTSFSK 9
| | | | |
Db 100 NQFYTAFSK 108

RESULT 8
US-09-141-821-3
; Sequence 3, Application US/09141821
; Patent No. 6110891
; GENERAL INFORMATION:
; APPLICANT: Arpad Janos PUSZTAI

```
; APPLICANT: Szuzanna Magdolna BARDOCZ
; APPLICANT: Richard Michael John PALMER
; APPLICANT: Neil William FISH
; APPLICANT: Gyorgy J. KOTELES
; TITLE OF INVENTION: LECTIN COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: 48656
; CURRENT APPLICATION NUMBER: US/09/141,821
; CURRENT FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Human
US-09-141-821-3
```

```
Query Match 66.7%; Score 30; DB 3; Length 274;
Best Local Similarity 85.7%; Pred. No. 74;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 SFMTSFS 8
Db 94 SFVTSFS 100
```

```
RESULT 9
US-09-141-821-1
; Sequence 1, Application US/09141821
; Patent No. 6110891
; GENERAL INFORMATION:
; APPLICANT: Arpad Janos PUSZTAI
; APPLICANT: Szuzanna Magdolna BARDOCZ
; APPLICANT: Richard Michael John PALMER
; APPLICANT: Neil William FISH
; APPLICANT: Gyorgy J. KOTELES
; TITLE OF INVENTION: LECTIN COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: 48656
; CURRENT APPLICATION NUMBER: US/09/141,821
; CURRENT FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Human
US-09-141-821-1
```

```
Query Match 66.7%; Score 30; DB 3; Length 285;
Best Local Similarity 85.7%; Pred. No. 77;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 SFMTSFS 8
Db 99 SFVTSFS 105
```

```
RESULT 10
US-09-141-821-2
; Sequence 2, Application US/09141821
; Patent No. 6110891
; GENERAL INFORMATION:
; APPLICANT: Arpad Janos PUSZTAI
; APPLICANT: Szuzanna Magdolna BARDOCZ
; APPLICANT: Richard Michael John PALMER
; APPLICANT: Neil William FISH
; APPLICANT: Gyorgy J. KOTELES
; TITLE OF INVENTION: LECTIN COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: 48656
; CURRENT APPLICATION NUMBER: US/09/141,821
; CURRENT FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 5
```

```
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Human
US-09-141-821-2
```

```
Query Match 66.7%; Score 30; DB 3; Length 285;
Best Local Similarity 85.7%; Pred. No. 77;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 SFMTSFS 8
Db 98 SFVTSFS 104
```

```
RESULT 11
US-09-141-821-4
; Sequence 4, Application US/09141821
; Patent No. 6110891
; GENERAL INFORMATION:
; APPLICANT: Arpad Janos PUSZTAI
; APPLICANT: Szuzanna Magdolna BARDOCZ
; APPLICANT: Richard Michael John PALMER
; APPLICANT: Neil William FISH
; APPLICANT: Gyorgy J. KOTELES
; TITLE OF INVENTION: LECTIN COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: 48656
; CURRENT APPLICATION NUMBER: US/09/141,821
; CURRENT FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Human
US-09-141-821-4
```

```
Query Match 66.7%; Score 30; DB 3; Length 285;
Best Local Similarity 85.7%; Pred. No. 77;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 SFMTSFS 8
Db 99 SFVTSFS 105
```

```
RESULT 12
US-09-141-821-5
; Sequence 5, Application US/09141821
; Patent No. 6110891
; GENERAL INFORMATION:
; APPLICANT: Arpad Janos PUSZTAI
; APPLICANT: Szuzanna Magdolna BARDOCZ
; APPLICANT: Richard Michael John PALMER
; APPLICANT: Neil William FISH
; APPLICANT: Gyorgy J. KOTELES
; TITLE OF INVENTION: LECTIN COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: 48656
; CURRENT APPLICATION NUMBER: US/09/141,821
; CURRENT FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Human
US-09-141-821-5
```

```
Query Match 66.7%; Score 30; DB 3; Length 285;
```

Best Local Similarity 85.7%; Pred. No. 77;
Matches 6; Conservative 1; Mismatches 0;

Indels 0; Gaps 0;

QY 2 SFMTSFS 8
||:||||
Db 99 SFVTSFS 105

RESULT 13
US-08-330-978-2
; Sequence 2, Application US/08330978
; Patent No. 5589571
; GENERAL INFORMATION:
; APPLICANT: King, Robert
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF INHIBITED
; TITLE OF INVENTION: FORMS OF ACTIVATED BLOOD FACTORS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/330,978
; FILING DATE: 28-OCT-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/484,558
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2803-0007.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)887-1500
; TELEFAX: (202)822-0168
; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 139 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 17..22
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 50..61
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 55..70
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 72..81
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 89..100
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 96..109
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 111..124
; FEATURE:
; NAME/KEY: Disulfide-bond

LOCATION: 132
OTHER INFORMATION: /note= "Disulfide linkage with
OTHER INFORMATION: residue 160 of SEQ ID NO:1, residue 108 of SEQ ID NO:3 or
OTHER INFORMATION: residue 108 of SEQ ID:4"
US-08-330-978-2

Query Match 64.4%; Score 29; DB 1; Length 139;
Best Local Similarity 55.6%; Pred. No. 56;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
||:|:|
Db 2 NSFLTMMK 10

RESULT 14
US-08-474-042-2
; Sequence 2, Application US/08474042
; Patent No. 5589572
; GENERAL INFORMATION:
; APPLICANT: King, Robert
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF INHIBITED
; TITLE OF INVENTION: FORMS OF ACTIVATED BLOOD FACTORS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,042
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/484,558
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2803-0007.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)887-1500
; TELEFAX: (202)822-0168
; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 139 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 17..22
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 50..61
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 55..70
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 72..81
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 89..100

FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 96..109
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 111..124
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 132
OTHER INFORMATION: /note= "Disulfide linkage with
OTHER INFORMATION: residue 160 of SEQ ID NO:1, residue 108 of SEQ ID NO:3 or
OTHER INFORMATION: residue 108 of SEQ ID:4"
US-08-474-042-2

Query Match 64.4%; Score 29; DB 1; Length 139;
Best Local Similarity 55.6%; Pred. No. 56;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 2 NSFLTTMKK 10

RESULT 15

US-08-484-558-2
Sequence 2, Application US/08484558
Patent No. 5602333
GENERAL INFORMATION:
APPLICANT: King, Robert
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF INHIBITED
TITLE OF INVENTION: FORMS OF ACTIVATED BLOOD FACTORS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,558
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2803-0007.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)887-1500
TELEFAX: (202)822-0168
TELEX: 90-4030 MRSNFOERSWSH
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 139 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 17..22
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 50..61
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 55..70
FEATURE:

NAME/KEY: Disulfide-bond
LOCATION: 72..81
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 89..100
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 96..109
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 111..124
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 132
OTHER INFORMATION: /note= "Disulfide linkage with
OTHER INFORMATION: residue 160 of SEQ ID NO:1, residue 108 of SEQ ID NO:3 or
OTHER INFORMATION: residue 108 of SEQ ID:4"
US-08-484-558-2

Query Match 64.4%; Score 29; DB 1; Length 139;
Best Local Similarity 55.6%; Pred. No. 56;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 2 NSFLTTMKK 10

Search completed: May 31, 2002, 10:27:51
Job time: 153 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 10:27:22 ; Search time 51.8 seconds
(without alignments)
19.299 Million cell updates/sec

Title: US-09-589-777C-25
Perfect score: 45
Sequence: 1 NSPMTSFSK 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	45	100.0	9	20	AA18408	Anti-angiogenic pe
2	45	100.0	50	22	AAB35587	Antiangiogenic pen
3	45	100.0	184	20	AA18409	Endostatin protein
4	45	100.0	184	20	AA18409	Murine endostatin
5	45	100.0	184	21	AA170258	Murine angiogenesi
6	45	100.0	184	22	AAB49380	Murine endostatin
7	45	100.0	191	21	AAB28398	Murine endostatin
8	45	100.0	207	22	AAE02031	Murine endostatin
9	45	100.0	207	22	AAB71930	Murine endostatin
10	45	100.0	218	20	AA18409	Murine gene therap
11	45	100.0	580	20	AA18409	Murine gene therap

12	45	100.0	1288	18	AAW26328	Mouse alpha-1 coll
13	42	93.3	185	20	AA106197	Anti-angiogenic en
14	40	88.9	1288	20	AAW22927	Mouse alpha-1 (XVI
15	38	84.4	184	21	AA170265	Canine angiogenesi
16	34	75.6	913	22	AAU45310	Propionibacterium
17	33	73.3	108	19	AA186022	S. pneumoniae deri
18	33	73.3	108	21	AA181753	Streptococcus pneu
19	32	71.1	60	22	AAW82463	Human immune/haema
20	32	71.1	246	19	AAW98254	H. pylori CHPO 149
21	32	71.1	395	12	AA110135	Fungal stress prot
22	32	71.1	593	19	AAW62835	Zea mays antimicro
23	31	68.9	53	21	AAB16814	Bacteriophage Dp-1
24	31	68.9	68	22	ABB03848	Human musculoskele
25	31	68.9	161	21	AAB43484	Human cancer assoc
26	31	68.9	236	16	AAW64807	PNA lectin subunit
27	31	68.9	314	22	AAG91752	C glutamicum prote
28	31	68.9	314	22	AAW79059	Corynebacterium gl
29	31	68.9	348	22	ABG05997	Novel human diagn
30	31	68.9	520	22	AAU18015	Human immunoglobul
31	31	68.9	697	22	AAW93829	Human polypeptide,
32	31	68.9	744	22	ABB10202	Human cDNA SEQ ID
33	31	68.9	801	22	AAG82463	S. epidermidis ope
34	31	68.9	1003	22	ABB68762	Drosophila melanog
35	31	68.9	1059	20	AA108095	Human PRO335 prote
36	31	68.9	1059	20	AA113393	Amino acid sequenc
37	31	68.9	1059	21	AA170672	Human PRO335 prote
38	31	68.9	1059	22	AAU00825	Human immune respo
39	31	68.9	1059	22	AAB80261	Human PRO335 prote
40	31	68.9	1119	20	AA108114	Human PRO326 prote
41	31	68.9	1119	21	AA170674	Human PRO326 prote
42	31	68.9	1119	22	AAU12347	Human PRO326 polyp
43	31	68.9	1119	22	AAU00827	Human immune respo
44	31	68.9	1119	22	AAB80263	Human PRO326 prote
45	31	68.9	1119	22	AAB48162	Human PRO326 polyp

ALIGNMENTS

RESULT 1

AA18408
ID AA18408 standard; peptide; 9 AA.

XX AA18408;

XX 24-AUG-1999 (first entry)

DE Anti-angiogenic peptide EMI deleted fragment.

KW EMI; anti-angiogenic peptide; endostatin; angiogenesis-dependent cancer;
KW benign tumour; rheumatoid arthritis; psoriasis; ocular angiogenesis;
KW Osler-Webber Syndrome; myocardial angiogenesis; angiofibroma; cancer;
KW plaque neovascularisation; telangiectasia; atherosclerosis; scleroderma;
KW dialysis graft vascular access stenosis; renal cancer; therapy.

XX Mus sp.

XX WO9929855-A1.

XX 17-JUN-1999.

XX 08-DEC-1998; 98WO-US26057.

XX 16-NOV-1998; 98US-0108536.

XX 08-DEC-1997; 97US-0067888.

XX 22-APR-1998; 98US-0082663.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Sukhatme VP;

XX WPI; 1999-385604/32.

XX

PT Mutant endostatin having anti-angiogenic activity
XX
PS Claim 4; Page 71; 105pp; English.
XX
CC This sequence is a fragment deleted from the mutant endostatin (EM)
CC of the invention, which has anti-angiogenic activity, and is designated
CC EMI. Compositions comprising EMI or fusion proteins comprising EMI, are
CC useful for treating diseases characterised by angiogenic activity, such
CC as angiogenesis-dependent cancers, benign tumours, rheumatoid arthritis,
CC psoriasis, ocular angiogenesis, Osler-Webber Syndrome, myocardial
CC angiogenesis, plaque neovascularisation, telangiectasia, haemophilic
CC joints, angiofibroma, wound granulation, intestinal adhesions,
CC atherosclerosis, scleroderma, hypertrophic scars, cat scratch disease,
CC Helicobacter pylori ulcers, dialysis graft vascular access stenosis,
CC contraception and obesity. In particular, the diseases treatable by EMI
CC comprise cancer, especially renal cancer. The methods provide a means for
CC introducing EMI into mammalian cells via gene therapy, for production of
CC EMI via recombinant means, as well as recombinant production of the EMI
CC protein. EMI performs as well or better than whole endostatin. Use
CC of EMI is advantageous for treatment of angiogenic diseases in that
CC increasingly smaller peptides are more potent on a weight basis, and may
CC be able to better penetrate tissues.
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 45; DB 20; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db | | | | | | | | | |
1 nsfmtsfsk 9

RESULT 2
AAB35587
ID AAB35587 standard; peptide; 50 AA.
XX
AC AAB35587;
XX

DT 14-FEB-2001 (first entry)
XX
DE Antiangiogenic pentacontapeptide IV.
XX
KW Antiangiogenic; angiogenesis; cancer; endostatin.
XX
OS Synthetic.
XX

PN WO200063249-A1.
XX
PD 26-OCT-2000.
XX
PF 11-APR-2000; 2000WO-EP03236.
XX
PR 15-APR-1999; 99IT-MI00777.
XX
PA (UYMI-) UNIV MILANO.
PA (UYFI-) UNIV FIRENZE.
XX
PI Chillemi F, Francescato P, Ziche M;
XX
DR WPI; 2001-007005/01.
XX

PT Polypeptides derived from endostatin exhibiting antiangiogenic activity
PT useful for treatment of angiogenesis-dependent tumours -
XX
PS Claim 5; Page 18; 28pp; English.
XX

CC The present invention describes a number of peptides derived from
CC endostatin which exhibit antiangiogenic activity. These may be used in
CC the treatment of cancer. The present sequence is one of the peptides of
CC the invention.

XX
SQ Sequence 50 AA;

Query Match 100.0%; Score 45; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.066;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db | | | | | | | | | |
42 nsfmtsfsk 50

RESULT 3
AAY18409
ID AAY18409 standard; Protein; 184 AA.
XX
AC AAY18409;
XX
DT 24-AUG-1999 (first entry)
XX
DE Endostatin protein sequence.
XX

KW EMI; anti-angiogenic peptide; endostatin; angiogenesis-dependent cancer;
KW benign tumour; rheumatoid arthritis; psoriasis; ocular angiogenesis;
KW Osler-Webber Syndrome; myocardial angiogenesis; angiofibroma; cancer;
KW plaque neovascularisation; telangiectasia; atherosclerosis; scleroderma;
KW dialysis graft vascular access stenosis; renal cancer; therapy.
XX

OS Mus sp.
XX
PN WO9929855-A1.
XX
PD 17-JUN-1999.
XX
PF 08-DEC-1998; 98WO-US26057.
XX
PR 16-NOV-1998; 98US-0108536.
PR 08-DEC-1997; 97US-0067888.
PR 22-APR-1998; 98US-0082663.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Sukhatme VP;
XX
DR WPI; 1999-385604/32.
DR N-PSDB; AAX79949.
XX

PT Mutant endostatin having anti-angiogenic activity
PS Claim 31; Fig 2; 105pp; English.
XX

CC This sequence is the mouse endostatin. The invention relates to a
CC the mutant endostatin (EM), which has anti-angiogenic activity, and is
CC designated EMI. Compositions comprising EMI or fusion proteins comprising
CC EMI, are useful for treating diseases characterised by angiogenic
CC activity, such as angiogenesis-dependent cancers, benign tumours,
CC rheumatoid arthritis, psoriasis, ocular angiogenesis, Osler-Webber
CC Syndrome, myocardial angiogenesis, plaque neovascularisation,
CC telangiectasia, haemophilic joints, angiofibroma, wound granulation,
CC intestinal adhesions, atherosclerosis, scleroderma, hypertrophic scars,
CC cat scratch disease, Helicobacter pylori ulcers, dialysis graft vascular
CC access stenosis, contraception and obesity. In particular, the diseases
CC treatable by EMI comprise cancer, especially renal cancer. The methods
CC provide a means for introducing EMI into mammalian cells via gene
CC therapy, for production of EMI via recombinant means, as well as
CC recombinant production of the EMI protein. EMI performs as well or better
CC than whole endostatin. Use of EMI is advantageous for treatment of
CC angiogenic diseases in that increasingly smaller peptides are more potent
CC on a weight basis, and may be able to better penetrate tissues.

XX Sequence 184 AA;
SQ

Query Match 100.0%; Score 45; DB 20; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
| | | | | | | |
Db 176 nsfmtsfsk 184

RESULT 4
AAY08689
ID AAY08689 standard; Protein; 184 AA.
XX
AC AAY08689;
XX
DT 10-AUG-1999 (first entry)
XX
DE Murine endostatin protein fragment.
XX
KW Plasminogen; murine; angiotatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina.
XX
OS Mus sp.
XX WO9926480-A1.
PN
XX 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-US24950.
XX
PR 20-NOV-1997; 97US-0975424.
XX
PA (GENE-) GENETIX PHARM INC.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Bachelot T, Leboulch P, Pawliuk RJ;
XX
DR WPI; 1999-357696/30.
DR N-PSDB; AAX77715.
XX
PT Anti-angiogenic gene therapy vectors
XX
PS Disclosure; Fig 6; 83pp; English.
XX
CC This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiotatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.
XX
SQ Sequence 184 AA;

Query Match 100.0%; Score 45; DB 20; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
| | | | | | | |
Db 176 nsfmtsfsk 184

RESULT 5
AAY70258

ID AAY70258 standard; Protein; 184 AA.
XX
AC AAY70258;
XX
DT 06-JUN-2000 (first entry)
XX
DE Murine angiogenesis inhibitor, endostatin.
XX
KW Murine; immunoglobulin Fc fragment; endostatin; immunofusin;
KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
KW vasotropic; vulnery; treatment; antiarteriosclerosis; tumour;
KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW wound granulation; keloid scar; gene therapy.
XX
OS Mus musculus.
XX
PN WO200011033-A2.
XX
PD 02-MAR-2000.
XX
PF 25-AUG-1999; 99WO-US19329.
XX
PR 25-AUG-1998; 98US-0097883.
XX
PA (LEXI-) LEXINGEN PHARM CORP.
XX
PI Lo K, Li Y, Gillies SD;
XX
DR WPI; 2000-237616/20.
DR N-PSDB; AA251299.
XX
PT Novel fusion protein of angiotatin or endostatin and an immunoglobulin
PT Fc region, useful for treating conditions mediated by angiogenesis,
PT such as rheumatoid arthritis, tumors and macular degeneration -
XX
PS Example 5; Pages 48-49; 68pp; English.
XX
CC The patent discloses a DNA molecule encoding a fusion protein comprising
CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis
CC inhibitor selected from angiotatin, endostatin, a plasminogen fragment
CC having angiotatin activity, a collagen XVIII fragment having endostatin
CC activity, or combinations of them. The fusion protein (immunofusin) is
CC used to inhibit angiogenesis and to treat diseases or conditions mediated
CC by angiogenesis. Conditions that may be treated include solid tumours,
CC blood born tumours, tumour metastasis, benign tumours including
CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
CC e.g. diabetic retinopathy, retinopathy of prematurity, macular
CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental
CC fibroplasia, rubecosis and Osler-Webber syndrome; myocardial angiogenesis,
CC plaque neovascularisation, telangiectasia, haemophilic joints,
CC angiofibroma, wound granulation, and excessive or abnormal stimulation of
CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used
CC in gene therapy. The present sequence is a murine
CC endostatin used in the construction of immunofusin containing murine
CC immunoglobulin Fc fragment.
XX
SQ Sequence 184 AA;

Query Match 100.0%; Score 45; DB 21; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
| | | | | | | |
Db 176 nsfmtsfsk 184

RESULT 6
AAB49380
ID AAB49380 standard; Protein; 184 AA.
XX
AC AAB49380;
XX
DT 02-MAR-2001 (first entry)
XX
DE Murine endostatin SEQ ID NO: 4.
XX
KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;
KW cancer; inflammation; angiogenesis-dependent disease.
XX
OS Mus musculus.
XX WO200067771-A1.
PN
XX 16-NOV-2000.
PD
XX 02-MAY-2000; 2000WO-US12063.
PF
XX 06-MAY-1999; 99US-0132907.
PR
PR 14-JUL-1999; 99US-0353333.
XX
PA (BURN-) BURNHAM INST.
XX
XX Vuori K;
PI
XX WPI; 2001-040937/05.
DR
DR N-PSDB; AAC88290.
XX
XX Endostatin peptide comprising at least four endostatin amino acid
PT residues are e.g. angiogenesis inhibitors for treating cancer and
PT diabetic retinopathy.
XX
XX Disclosure; Fig 1; 146pp; English.
PS
XX
XX The present invention provides endostatin peptides which can be used in
CC the modulation of angiogenesis. This is useful in the treatment of
CC cancers, inflammation, rheumatoid arthritis, chronic articular
CC rheumatism, psoriasis, disorders associated with inopportune invasion of
CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
CC of prematurity, macular degeneration, corneal graft rejection,
CC retrolental fibroplasia, rubeosis, capillary proliferation in
CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
CC diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophilic joints and wound
CC granulation. In addition, the peptides can be used as birth control
CC agents.
XX
XX Sequence 184 AA;
SQ

Query Match 100.0%; Score 45; DB 22; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 176 nsfmtsfsk 184
|||||

RESULT 7
AAB28398
ID AAB28398 standard; Protein; 191 AA.
XX
AC AAB28398;
XX
DT 19-FEB-2001 (first entry)
XX
DE Murine endostatin.
XX
XX Murine; endostatin; cytostatic; antiproliferative;

KW vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;
KW cancer; vascularised solid tumour.
XX
OS Mus sp.
PN WO200064946-A2.
XX
PD 02-NOV-2000.
XX
PF 28-APR-2000; 2000WO-US11367.
XX
PR 28-APR-1999; 99US-0131432.
XX
XX (TEXA) UNIV TEXAS SYSTEM.
XX
XX Thorpe PE, Brekken RA;
PI
XX WPI; 2000-687317/67.
DR
DR N-PSDB; AAC67777.
XX
XX Immunogenic composition for the treatment and diagnosis of cancer
PT comprises an anti-VEGF (vascular endothelial growth factor) antibody
PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -
XX
XX Example 10; Page 290-291; 298pp; English.
PS
XX
XX The present invention relates to anti-Vascular Endothelial Growth Factor
CC (VEGF) antibodies that bind to the same epitope as the monoclonal
CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to
CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF
CC receptor VEGFR1. The present sequence is murine endostatin. Endostatin
CC may be conjugated onto the anti-VEGF antibodies of the present invention.
CC The anti-VEGF antibodies of the present invention are useful for the
CC treatment and diagnosis of cancer, especially vascularised solid tumours.
XX
XX Sequence 191 AA;
SQ

Query Match 100.0%; Score 45; DB 21; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 183 nsfmtsfsk 191
|||||

RESULT 8
AAE02031
ID AAE02031 standard; Protein; 207 AA.
XX
AC AAE02031;
XX
DT 31-JUL-2001 (first entry)
XX
DE Murine endostatin fused to N-terminal secretion signal.
XX
KW Murine; endostatin; fusion protein; nucleotide-binding domain; NBD;
KW ligand-binding domain; LBD; transcription regulating domain; TRD; cancer;
KW zinc finger protein; ZFP; ligand-activated transcriptional regulator;
KW gene regulation; gene therapy; cell proliferative disorder; psoriasis;
KW pemphigus vulgaris; Behcet's syndrome; lipid histiocytosis.
XX
XX Mus sp.
OS
XX
XX WO200130843-A1.
PN
XX
XX 03-MAY-2001.
PD
XX
XX 23-OCT-2000; 2000WO-EP10430.
PF
XX 25-OCT-1999; 99US-0433042.
PR
PR 02-JUN-2000; 2000US-0586625.
PR

XX (NOVS) NOVARTIS AG.
PA (SCRI) SCRIPPS RES INST.
XX
PI Barbas CF, Kadan M, Beerli R;
XX
DR WPI; 2001-308618/32.
DR N-PSDB; AAD06108.
XX
PT New fusion protein containing nucleotide-binding and ligand-binding
PT domains, useful e.g. in gene therapy of cancer, provides
PT ligand-activated control of gene expression -
XX
PS Example 19; Page 209; 218pp; English.
XX
CC The invention relates to fusion protein comprising a nucleotide-binding
CC domain (NBD), a ligand-binding domain (LBD) of an intracellular receptor
CC (ICR) and a transcription regulating domain (TRD). NBD is a polydactyl
CC zinc finger protein (ZFP), or a modular part of it, that interacts
CC specifically with a contiguous sequence of at least 3 nucleotides. The
CC fusion protein functions as a ligand-activated transcriptional regulator.
CC The fusion protein and the nucleic acid encoding it, are used to regulate
CC gene expression, particularly in gene therapy for treating malignant
CC cell proliferative diseases (e.g. colon cancer, prostate cancer,
CC renal-cell carcinoma) and non-malignant cell proliferative
CC diseases (e.g. psoriasis, pemphigus vulgaris, Behcet's syndrome and
CC lipid histiocytosis). The fusion protein and its DNA are also useful for
CC treating diseases caused by viruses in humans/plants, genetic and/or
CC acquired diseases. The fusion protein can be designed to target any
CC selected gene (endogenous or exogenous), and can be made to have
CC different selectivity or specificity for endogenous or exogenous ligands.
CC The present sequence is murine endostatin fused to an N-terminal
CC secretion signal. The corresponding cDNA sequence was used in the
CC construction of Left end shuttle plasmids containing regulatable
CC transgene cassettes for evaluation of Cys2-His2 Zinc finger DNA binding
CC domain (DBD)-Oestrogen receptor (ER) LBD regulators.
XX
SQ Sequence 207 AA;

Query Match 100.0%; Score 45; DB 22; Length 207;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||
Db 199 nsfmtsfsk 207

RESULT 9
AAB71930
ID AAB71930 standard; Protein; 207 AA.
XX
AC AAB71930;
XX
DT 10-MAY-2001 (first entry)
XX
DE Murine endostatin attached to Ig-kappa leader sequence.
XX
KW Mouse; endostatin; antitumour; cytostatic; antiarthritic; antipsoriatic;
KW antidiabetic; ophthalmological; gene therapy; angiogenic inhibitor;
KW adenoviral vector; diabetic retinopathy; cardiovascular disease;
KW arthritis; psoriasis; cerebral oedema; intravascular coagulopathy;
KW lymphoma; leukaemia; immunoglobulin; Ig; Ig-kappa.
XX
OS Mus sp.
XX
PN WO200112830-A1.
XX
PD 22-FEB-2001.
XX
PF 11-AUG-2000; 2000WO-EP07865.
XX

PR 13-AUG-1999; 99US-0373938.
XX
PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Hallenbeck PL, Chen CT;
XX
DR WPI; 2001-202871/20.
DR N-PSDB; AAF60336.
XX
PT Adenoviral vector for treating tumors and disorders associated with
PT angiogenesis, such as cancer, arthritis, and psoriasis, comprises a DNA
PT sequence encoding an angiogenic inhibitor, particularly endostatin -
XX
PS Example 1; Fig 1B; 59pp; English.
XX
CC The nucleotide sequence encoding this protein was used in the
CC construction of an adenoviral vector which includes a DNA sequence
CC encoding endostatin. The adenoviral vector is useful for expressing
CC endostatin in a mammalian cell such as an A549 or Hep3B cell. It is
CC useful for treating other diseases and disorders associated with
CC angiogenesis, such as neovascular diseases of the eye, including diabetic
CC retinopathy, cardiovascular disease, arthritis, psoriasis, cerebral
CC oedema and intravascular coagulopathy (Kasabach-Merritt syndrome). The
CC vector inhibits, prevents or destroys the growth of tumours by
CC preventing the formation of blood vessels in tumours, such as lymphoma
CC and leukaemia.
XX
SQ Sequence 207 AA;

Query Match 100.0%; Score 45; DB 22; Length 207;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
|||||
Db 199 nsfmtsfsk 207

RESULT 10
AAY08691
ID AAY08691 standard; Protein; 218 AA.
XX
AC AAY08691;
XX
DT 10-AUG-1999 (first entry)
XX
DE Murine gene therapy peptide construct SP-Flag-Endo.
XX
KW plasminogen; murine; angiostatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.
XX
OS Mus sp.
OS Synthetic.
XX
PN WO9926480-A1.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-US24950.
XX
PR 20-NOV-1997; 97US-0975424.
XX
PA (GENE-) GENETIX PHARM INC.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Bachelot T, Leboulch P, Pawliuk RJ;
XX
DR WPI; 1999-357696/30.
DR N-PSDB; AAX77717.
XX

PT Anti-angiogenic gene therapy vectors
XX
PS Example 1; Page 69; 83pp; English.
XX
CC This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiotatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.
XX
SQ Sequence 218 AA;

Query Match 100.0%; Score 45; DB 20; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 210 nsfmtsfsk 218
|||||

RESULT 11
AAY08692
ID AAY08692 standard; Protein; 580 AA.
XX
AC AAY08692;
XX
DT 10-AUG-1999 (first entry)
XX
DE Murine gene therapy peptide construct SP-K1-K2-K3-K4-Flag-Endo.
XX
KW Plasminogen; murine; angiotatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.
XX
OS Mus sp.
OS Synthetic.
XX
PN WO9926480-A1.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-US24950.
XX
PR 20-NOV-1997; 97US-0975424.
XX
PA (GENE-) GENETIX PHARM INC.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Bachelot T, Leboulch P, Pawliuk RJ;
XX
DR WPI; 1999-357696/30.
DR N-PSDB; AAX77718.
XX
PT Anti-angiogenic gene therapy vectors
XX
PS Example 1; Page 72-74; 83pp; English.
XX

CC This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiotatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for

CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.
XX
SQ Sequence 580 AA;

Query Match 100.0%; Score 45; DB 20; Length 580;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
Db 572 nsfmtsfsk 580
|||||

RESULT 12
AAW26328
ID AAW26328 standard; Protein; 1288 AA.
XX
AC AAW26328;
XX
DT 19-NOV-1997 (first entry)
XX
DE Mouse alpha-1 collagen (XVIII).
XX
KW Alpha-1 collagen; type XVIII collagen; cartilage degeneration.
XX
OS Mus musculus.

Key	Peptide	Location/Qualifiers
FT	Peptide	303..308
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	309..314
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	315..320
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	321..326
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	337..342
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	343..348
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	349..354
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	355..360
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	361..366
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	367..372
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	373..378
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	379..384
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	385..390
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	396..401
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	402..407
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	435..440
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	441..446
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	447..452
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	453..458
FT	Peptide	/label= GXYGX'Y'_motif
FT	Peptide	459..464
FT	Peptide	/label= GXYGX'Y'_motif

FT Peptide 470..475 /label= GXYGX'Y'_motif
FT Peptide 476..481 /label= GXYGX'Y'_motif
FT Peptide 482..487 /label= GXYGX'Y'_motif
FT Peptide 488..493 /label= GXYGX'Y'_motif
FT Peptide 494..499 /label= GXYGX'Y'_motif
FT Peptide 500..505 /label= GXYGX'Y'_motif
FT Peptide 506..511 /label= GXYGX'Y'_motif
FT Peptide 512..517 /label= GXYGX'Y'_motif
FT Peptide 518..523 /label= GXYGX'Y'_motif
FT Peptide 524..529 /label= GXYGX'Y'_motif
FT Peptide 530..535 /label= GXYGX'Y'_motif
FT Peptide 536..541 /label= GXYGX'Y'_motif
FT Peptide 542..547 /label= GXYGX'Y'_motif
FT Peptide 548..553 /label= GXYGX'Y'_motif
FT Peptide 580..585 /label= GXYGX'Y'_motif
FT Peptide 586..591 /label= GXYGX'Y'_motif
FT Peptide 592..597 /label= GXYGX'Y'_motif
FT Peptide 598..603 /label= GXYGX'Y'_motif
FT Peptide 604..609 /label= GXYGX'Y'_motif
FT Peptide 610..615 /label= GXYGX'Y'_motif
FT Peptide 616..621 /label= GXYGX'Y'_motif
FT Peptide 622..627 /label= GXYGX'Y'_motif
FT Peptide 628..633 /label= GXYGX'Y'_motif
FT Peptide 634..639 /label= GXYGX'Y'_motif
FT Peptide 640..665 /label= GXYGX'Y'_motif
FT Peptide 657..662 /label= GXYGX'Y'_motif
FT Peptide 677..682 /label= GXYGX'Y'_motif
FT Peptide 683..688 /label= GXYGX'Y'_motif
FT Peptide 689..694 /label= GXYGX'Y'_motif
FT Peptide 695..700 /label= GXYGX'Y'_motif
FT Peptide 707..712 /label= GXYGX'Y'_motif
FT Peptide 713..718 /label= GXYGX'Y'_motif
FT Peptide 735..740 /label= GXYGX'Y'_motif
FT Peptide 741..746 /label= GXYGX'Y'_motif
FT Peptide 747..752 /label= GXYGX'Y'_motif
FT Peptide 759..764 /label= GXYGX'Y'_motif
FT Peptide 765..770 /label= GXYGX'Y'_motif

FT Peptide /label= GXYGX'Y'_motif
FT Peptide 771..776 /label= GXYGX'Y'_motif
FT Peptide 787..792 /label= GXYGX'Y'_motif
FT Peptide 793..798 /label= GXYGX'Y'_motif
FT Peptide 799..804 /label= GXYGX'Y'_motif
FT Peptide 815..820 /label= GXYGX'Y'_motif
FT Peptide 821..826 /label= GXYGX'Y'_motif
FT Peptide 827..832 /label= GXYGX'Y'_motif
FT Peptide 833..838 /label= GXYGX'Y'_motif
FT Peptide 839..844 /label= GXYGX'Y'_motif
FT Peptide 845..850 /label= GXYGX'Y'_motif
FT Peptide 863..868 /label= GXYGX'Y'_motif
FT Peptide 869..874 /label= GXYGX'Y'_motif
FT Peptide 875..880 /label= GXYGX'Y'_motif
FT Peptide 891..896 /label= GXYGX'Y'_motif
FT Peptide 897..902 /label= GXYGX'Y'_motif
FT Peptide 903..908 /label= GXYGX'Y'_motif
FT Peptide 911..916 /label= GXYGX'Y'_motif
FT Peptide 917..922 /label= GXYGX'Y'_motif
FT Peptide 928..933 /label= GXYGX'Y'_motif
FT Peptide 934..939 /label= GXYGX'Y'_motif
FT Peptide 956..961 /label= GXYGX'Y'_motif
FT Peptide 962..967 /label= GXYGX'Y'_motif
FT Peptide 968..973 /label= GXYGX'Y'_motif
FT Peptide 1126..1131 /label= GXYGX'Y'_motif
FT Peptide 1145..1150 /label= GXYGX'Y'_motif
FT Peptide 1193..1198 /label= GXYGX'Y'_motif
XX US5643783-A.
PN 01-JUL-1997.
XX 01-DEC-1993; 93US-0159784.
XX 01-DEC-1993; 93US-0159784.
XX (HARD) HARVARD COLLEGE.
PI Oh SP, Olsen BR;
XX WPI; 1997-350247/32.
DR N-PSDB; AAT84485.
XX Nucleic acid encoding human alpha-1 collagen - for production of
PT recombinant alpha-1 collagen, for use in the treatment of cartilage
PT degeneration
XX

PS Disclosure; Fig 2; 35pp; English.

Query Match 100.0%; Score 45; DB 18; Length 1288;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMSTPSK 9
| | | | | | | | | |
Db 1280 nsfmstfsk 1288

RESULT 13
AAAY06197
ID AAY06197 standard; Protein; 185 AA.

XX AC AAY06197;
XX DT 16-AUG-1999 (first entry)
XX DE Anti-angiogenic endostatin peptide.
XX KW Anti-angiogenic; carrier:DNA complex; tumour; gene therapy; human;
XX KW endostatin; melanoma; lung cancer; colon cancer; brain cancer;
XX KW breast cancer.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Misc-difference 36 /note= "encoded by CAG"
FT Misc-difference 37 /note= "encoded by CAA"
FT Misc-difference 39 /note= "deduced sequence from nucleotide sequence
has an Ala residue between residues 39 and
40 of this sequence"
FT Misc-difference 76 /note= "encoded by AAG"
FT Misc-difference 118 /note= "encoded by AAG"
FT Misc-difference 162 /note= "encoded by AAA"
FT Misc-difference 168 /note= "encoded by AAC"
FT Misc-difference 185 /note= "encoded by AAA"
XX EP921193-A1.
XX 09-JUN-1999.
XX 07-JAN-1998; 98EP-0100135.
XX 05-DEC-1997; 97US-0985526.
XX (MIXS/) MIXSON A J.
XX Mixson AJ;
XX WPI; 1999-315406/27.
XX N-PSDB; AAX58740.
XX Inhibition of growth of solid tumors
XX Disclosure; Page 38; 46pp; English.
XX The present sequence represents an anti-angiogenic endostatin
peptide. The invention provides a carrier:DNA complex that comprises
DNA (see AAX58725-42) encoding an anti-angiogenic protein or peptide,
such as the present sequence, the complex being deliverable to
the site of a tumour in vivo, and additionally comprises regulatory
elements for expressing the anti-angiogenic DNA in a tumour or
tumour vasculature. The complex may also include DNA encoding a

CC tumour suppressor protein, especially p53. The carrier is a
CC liposome, cationic polymer, micelle, microsphere, virus, viral
CC component, or a combination of these, and administration is by
CC intravenous or intratumoral injection. The complexes are useful in
CC gene therapy for inhibition of tumour growth. The types of tumors
CC which may be treated include solid tumors such as melanomas and
CC tumors in the lung, colon, brain and breast.
XX Sequence 185 AA;
SQ

Query Match 93.3%; Score 42; DB 20; Length 185;
Best Local Similarity 88.9%; Pred. No. 0.93;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMSTPSK 9
| | | | | | | | | |
Db 177 nsfmstfsr 185

RESULT 14
AAW92297
ID AAW92297 standard; peptide; 1288 AA.
XX AC AAW92297;
XX DT 28-APR-1999 (first entry)
XX DE Mouse alpha-1 (XVIII) collagen chain common sequence MO18(common)28.
XX KW Human; type XVIII collagen; liver disease; cirrhosis; detection;
XX KW hepatocellular carcinoma; diagnosis.
XX OS Mus sp.
XX WO9856399-A1.
XX 17-DEC-1998.
XX 12-JUN-1998; 98WO-US12327.
XX 12-JUN-1997; 97US-0049369.
XX (FIFI-) ACAD FINLAND.
XX (FIBR-) FIBROGEN INC.
XX (INRM) INST NAT SANTE & RECH MEDICALE.
XX Clement B, Pihlajaniemi T, Rehn M;
XX WPI; 1999-070292/06.
XX Diagnosis and monitoring of liver disease by measuring collagen type
XX XVIII levels - with elevated levels indicative of disease,
XX especially cirrhosis or hepatocellular carcinoma
XX Example 6; Fig 8; 56pp; English.
XX A method has been developed for the detecting liver disease. The method
XX comprises: (a) reacting a patient sample with antibodies (Ab) specific
XX for collagen type XVIII (Coll18); (b) measuring the amount of Ab-antigen
XX complex (C) formed as indicator of the amount of Coll18 present; (c)
XX similar analysis of a non-diseased control; and (d) comparing the
XX amounts of Coll18 in the two samples to detect presence or progression of
XX disease. Elevated levels of Coll18 are: (i) indicative of disease,
XX specifically cirrhosis; and (ii) predictive of the prognosis of disease,
XX specifically hepatocellular carcinoma (there is a relationship between
XX Coll18 mRNA levels and tumour size and necrosis, and survival times are
XX significantly higher in patients with higher Coll18 levels). The method
XX provides non-invasive, early and accurate diagnosis of liver disease.
XX The present sequence represents the sequence common to mouse alpha-1
XX (XVIII) collagen chain from the present invention.
XX Sequence 1288 AA;
SQ

Query Match 88.9%; Score 40; DB 20; Length 1288;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NSFMTSPS 8
 |||||||
Db 1281 nsfmtsfs 1288

RESULT 15
AAAY70265
ID AAAY70265 standard; Protein; 184 AA.
XX
AC AAAY70265;
XX
DT 06-JUN-2000 (first entry)
XX
DE Canine angiogenesis inhibitor, endostatin.
XX
KW Canine; immunoglobulin Fc fragment; endostatin; immunofusin;
KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
KW vasotropic; vulnery; treatment; antiarteriosclerosis; tumour;
KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW wound granulation; keloid scar; gene therapy.

XX Canis familiaris.
OS
XX
XX WO200011033-A2.
PN
XX
XX 02-MAR-2000.
PD
XX
XX 25-AUG-1999; 99WO-US19329.
PF
XX
XX 25-AUG-1998; 98US-0097883.
PR
XX
XX (LEXI-) LEXINGEN PHARM CORP.
PA
XX
XX LO K, Li Y, Gillies SD;
PI
XX
DR WPI; 2000-237616/20.
DR N-PSDB; AA251309.

XX
PT Novel fusion protein of angiostatin or endostatin and an immunoglobulin
PT FC region, useful for treating conditions mediated by angiogenesis,
PT such as rheumatoid arthritis, tumors and macular degeneration -

XX Example 8; Pages 59-60; 68pp; English.

XX
CC The patent discloses a DNA molecule encoding a fusion protein comprising
CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis
CC inhibitor selected from angiostatin, endostatin, a plasminogen fragment
CC having angiostatin activity, a collagen XVIII fragment having endostatin
CC activity, or combinations of them. The fusion protein (immunofusin) is
CC used to inhibit angiogenesis and to treat diseases or conditions mediated
CC by angiogenesis. Conditions that may be treated include solid tumours,
CC blood born tumours, tumour metastasis, benign tumours including
CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
CC e.g. diabetic retinopathy, retinopathy of prematurity, macular
CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental
CC fibroplasia, rubeosis and Osler-Webber syndrome; myocardial angiogenesis,
CC plaque neovascularisation, telangiectasia, haemophilic joints,
CC angiofibroma, wound granulation, and excessive or abnormal stimulation of
CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used
CC in gene therapy. The present sequence is a canine
CC endostatin used in the construction of immunofusin containing canine
CC immunoglobulin Fc fragment.

XX
SQ Sequence 184 AA;

Query Match 84.4%; Score 38; DB 21; Length 184;
Best Local Similarity 88.9%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NSFMTSFSK 9
 |||||||
Db 176 nsvmtsfsk 184

Search completed: May 31, 2002, 10:27:23
Job time: 240 sec